Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Biomarkers selection for population normalization in SARS-CoV-2 wastewater-based epidemiology

Shu-Yu Hsu\textsuperscript{a, b}, Mohamed Bayati\textsuperscript{a}, Chenhui Li\textsuperscript{a}, Hsin-Yeh Hsieh\textsuperscript{a}, Anthony Belenchia\textsuperscript{c}, Jessica Klutts\textsuperscript{d}, Sally A. Zemmer\textsuperscript{d}, Melissa Reynolds\textsuperscript{c}, Elizabeth Semkiw\textsuperscript{c}, Hwei-Yiing Johnson\textsuperscript{c}, Trevor Foley\textsuperscript{e}, Chris G. Wieberg\textsuperscript{d}, Jeff Wenzel\textsuperscript{c}, Marc C. Johnson\textsuperscript{f}, Chung-Ho Lin\textsuperscript{a, b, *}

\textsuperscript{a} School of Natural Resources, University of Missouri, Columbia, MO 65201, USA
\textsuperscript{b} Center for Agroforestry, University of Missouri, Columbia, MO 65201, USA
\textsuperscript{c} Bureau of Environmental Epidemiology, Division of Community and Public Health, Missouri Department of Health and Senior Services, Jefferson City, MO, USA
\textsuperscript{d} Water Protection Program, Missouri Department of Natural Resources, Jefferson City, MO, USA
\textsuperscript{e} Missouri Department of Corrections, Jefferson City, MO, USA
\textsuperscript{f} Department of Molecular Microbiology and Immunology, University of Missouri, School of Medicine and the Christopher S. Bond Life Sciences Center, Columbia, MO 65201, USA

\begin{abstract}
Wastewater-based epidemiology (WBE) has been one of the most cost-effective approaches to track the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) levels in the communities since the coronavirus disease 2019 (COVID-19) outbreak in 2020. Normalizing SARS-CoV-2 concentrations by the population biomarkers in wastewater is critical for interpreting the viral loads, comparing the epidemiological trends among the sewersheds, and identifying the vulnerable communities. In this study, five population biomarkers, pepper mild mottle virus (PMMoV), creatinine (CRE), 5-hydroxyindoleacetic acid (5-HIAA), caffeine (CAF) and its metabolite paraxanthine (PARA) were investigated and validated for their utility in normalizing the SARS-CoV-2 loads through two normalizing approaches using the data from 64 wastewater treatment plants (WWTPs) in Missouri. Their utility in assessing the real-time population contributing to the wastewater was also evaluated. The best performing candidate was further tested for its capacity for improving correlation between normalized SARS-CoV-2 loads and the clinical cases reported in the City of Columbia, Missouri, a university town with a constantly fluctuating population. Our results showed that, except CRE, the direct and indirect normalization approaches using biomarkers allow accounting for the changes in wastewater dilution and differences in relative human waste input over time regardless flow volume and population of the given WWTP. Among selected biomarkers, PARA is the most reliable population biomarker in determining the SARS-CoV-2 load per capita due to its high accuracy, low variability, and high temporal consistency to reflect the change in population dynamics and dilution in wastewater. It also demonstrated its excellent utility for real-time assessment of the population contributing to the wastewater. In addition, the viral loads normalized by the PARA-estimated population significantly improved the correlation ($\rho=0.5878, p < 0.05$) between SARS-CoV-2 load per capita and case numbers per capita. This chemical biomarker complements the current normalization scheme recommended by CDC and helps us understand the size, distribution, and dynamics of local populations for forecasting the prevalence of SARS-CoV-2 within each sewershed.
\end{abstract}

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has caused a global pandemic declared by the World Health Organization (WHO) on March 11th, 2020 (\textit{WHO}, n.d.). Despite clinical tests being sufficient and accurate, their time-consuming and often expensive process has not always been sufficient enough to track SARS-CoV-2 outbreaks at the population scale (\textit{Gonzalez et al.}, 2020). Wastewater-based
epidemiology (WBE) offers near real-time information about the outbreak to track SARS-CoV-2 in the communities (Polo et al., 2020). It has been successfully used to predict the overall status of infection and to capture asymptomatic and pre-symptomatic infections in the given wastewater treatment plant (WWTP) served area (Agrawal et al., 2021). Several studies in Europe, Australia, Japan, Singapore and the United States had used WBE approach. (Agrawal et al., 2021; Ahmed et al., 2020; Haramoto et al., 2020; La Rosa et al., 2020; Medema et al., 2020; Randazzo et al., 2020; Shcheran et al., 2020; Wong et al., 2021; Wu et al., 2020). The State of Missouri launched a statewide wastewater SARS-CoV-2 surveillance program in May 2020. (State of Missouri, 2021). It has been successfully applied to (1) provide the early warning, (2) determine the distribution of SARS-CoV-2 and its variants in Missouri, (3) identify trends in SARS-CoV-2 prevalence in areas surveilled, and (4) monitor for indicators of SARS-CoV-2 reemergence to inform mitigation efforts.

For long-term wastewater SARS-CoV-2 surveillance, normalizing SARS-CoV-2 wastewater concentrations prior to calculating trends is recommended by the United States Centers for Disease Control (CDC) and the European Commission to account for changes in wastewater dilution and differences in relative human waste input over time, due to tourism, weekday commuters, temporary workers, and pandemic lockdowns, etc. Normalizing SARS-CoV-2 concentrations by the amount of human excreta in wastewater can be crucial for interpreting and comparing viral concentrations in the sewage samples over time (CDC, 2021a; European Commission, 2021).

The population biomarkers include but are not limited to viral or bacterial molecular targets; cross-assembly phase or pepper mild mottle virus (PMMoV) are recommended by both CDC and European Commission (CDC, 2021b; European Commission, 2021). PMMoV, a viral pathogen in Capsicum sp. that had been identified in several pepper-based products and diets (Zhang et al., 2005), is one of the fecal indicator viral molecular targets (Genda et al., 2005). PMMoV was utilized to track human fecal sources due to its abundance in pepper-based food, unaffected by seasonal change, persistence in the wastewater (with half-life from 6-10 days) from the populated area (Kitajima et al., 2014; Rosario et al., 2009). In recent studies, PMMoV was utilized for SARS-CoV-2 normalization of external factors such as flow rate of the wastewater (Jafferai et al., 2021) and background noise associated with systematic variation (D’Aoust et al., 2021). However, these viral or bacterial molecular have not been used to estimate the population contributing to the wastewater yet.

On the other hand, the chemical population biomarkers, specific to human and majorly excreted through feces and/or urine contributing to wastewater, had been utilized to estimate the size of the population (Chen et al., 2014; Driver et al., 2020; Gracia-Lor et al., 2017; O’Brien et al., 2017, 2014; Polo et al., 2020; Rico et al., 2017; Senta et al., 2015). Several chemical markers, such as creatinine (CRE), 5-hydroxyindoleacetic acid (5-HIAA), caffeine (CAF), and its metabolite paraxanthine (PARA) have been reported as promising candidates (Chen et al., 2014; Driver et al., 2020; O’Brien et al., 2017, 2014; Rico et al., 2017; Senta et al., 2015). Creatinine is the metabolite of creatine and phosphorlycreatine in the muscles. It is produced at a steady state, diffused out of muscle cells, and further excreted by kidneys into urine (Wyss and Kaddurah-Daouk, 2000). Urinary CRE was routinely used to account for dilution when testing human urine for illicit substances (Bart Dana et al., 2005; Brewer et al., 2012). The serotonin metabolite 5-HIAA is another promising chemical molecule for this purpose. Clinical urinary 5-HIAA analysis is commonly performed to evaluate patients with suspected carcinoid syndrome (Nuttall and Pintegre, 1998). The 5-HIAA been also used to estimate the population contributing to the wastewater (Rico et al., 2017). Both CRE and 5-HIAA had been quantified in the samples from WWTPs with the modern liquid-chromatography mass spectrometry (Chiaia et al., 2008; Jang et al., 2019). Ricoet al. (2017) reported that 5-HIAA loads in the WWTP samples showed a positive correlation with the population calculated using the hydrochemical parameters. Chen et al. (2014) reported that 5-HIAA levels were also correlated well with the census population.

Besides endogenous molecules, CAF, a widely consumed central nervous system (CNS) stimulant (dePaula and Farah, 2019), is commonly found in food products, including tea, coffee, energy drinks and chocolate, as well as in some medications and dietary supplements. The PARA is the major metabolite of CAF through the cytochrome P4501A2 (CYP1A2)-catalyzed 3-demethylation (Crews et al., 2010). Several studies had detected CAF and PARA in the wastewater (Chiaia et al., 2008; Driver et al., 2020; Gracia-Lor et al., 2017; O’Brien et al., 2014; Senta et al., 2015). Similar to 5-HIAA, research have reported a positive correlation between CAF load and the population from census or population calculated by the hydrochemical parameters (O’Brien et al., 2014; Rico et al., 2017). The PARA level was found less affected by the genetic heterogeneity and population structure as compared to its parent compound CAF (Crews et al., 2010), suggesting PARA could also be a potential population biomarker.

The goal of this study was to determine the most suitable population biomarker for SARS-CoV-2 wastewater surveillance. Specific objectives were (1) to compare the variability and accuracy of the selected biomarkers for normalizing the SARS-CoV2 concentrations using two developed approaches derived from biomarker concentration, (2) to identify the reliable biomarkers for estimating the real-time population contributing to the wastewater, and (3) to demonstrate the SARS-CoV-2 loads normalized with the selected biomarkers against COVID-19 cases reported by the local health care authority among residents of the sewershed.

2. Material and method

2.1. Wastewater sampling

To develop the relationship between biomarkers and population, triplicates of 50 mL of the 24-h composite wastewater samples were collected once per week from the raw inlets, before the primary treatment, at 12 WWTPs (Table 1) in Missouri from 18th to 29th in January 2021.

Following the correlation analysis, wastewater composite samples collected from 64 WWTPs (Table S1) across the State of Missouri, were used for method validation. They were collected during the week of May 10th in 2021. The WWTPs serve urban, semirural, and rural locations throughout Missouri with the sewershed population ranging from 4600 to 306,647 (number of people estimated by WWTPs or Missouri Census). Ten wastewater composite samples were collected from WWTPs at the City of Columbia (college town) and a tourist town respectively through May to early September in 2021 (Table S2) for evaluating the utility of the biomarker for assessing the population fluctuation and dynamics. All of wastewater samples were transported in coolers with cold packs and then stored at 4°C until further extraction within two days.

2.2. Quantification of SARS-CoV-2

2.2.1. RNA extraction from wastewater samples

Fifty mL of wastewater from each catchment was filtered through a 0.22-micron filter (Millipore cat#: SGPOO525). Thirty-six mL of filtered wastewater were mixed with 12 mL of 50% (W/V) polyethylene glycol (PEG, Research Products International, cat#: P48080) and 1.2 M NaCl for a final concentration of 12% PEG and 0.3 M NaCl, followed by incubation for 2 h at 4°C. Samples were further centrifuged at 12,000 Xg for 2 h. RNA was extracted from the pellet using Qiagen Viral RNA extraction kit following the manufacturer’s instructions after the supernatant was removed. RNA was eluted in a final volume of 60 µL (Robinson et al., 2022). The samples were stored at -80°C if not processed immediately.
3 alternative primer and probe sets. Final RT-qPCR one-step mixtures for the CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel (Acceptable Alternative Primer and Probe Sets) were chosen based on the sequences of PMMoV gene fragment, primers, and probes. Table 1 provides the site summary of the 12 wastewater treatment plants for the model development.

### Table 1

| No. | Project ID | City          | County        | Samples/Week | Population Served | Source of Population | \(\text{Capacity}^a\) | \(\text{Composite sampling mode}^b\) | \(\text{Daily influent flow}^c\) | \(\text{Daily influent flow}^c\) |
|-----|------------|---------------|---------------|--------------|-------------------|----------------------|----------------|------------------|------------------|------------------|
| 1   | CARTH     | Carthage      | Jasper        | 1            | 12000             | Operator information | 26.50          | Time Based      | 14.95            | 15.82            |
| 2   | WARNE     | Warrensburg   | Johnson       | 1            | 7990              | Operator information | 5.68           | Flow Based      | 3.40             | 3.19             |
| 3   | FULTN     | Fulton        | Callaway      | 1            | 12790             | Operator information | 10.98          | Time Based      | 6.06             | 13.25            |
| 4   | SFDNW     | Springfield   | Greene        | 1            | 26078             | Connections with population correction | 25.74          | Time Based      | 15.79            | 15.90            |
| 5   | HANBL     | Hannibal      | Marion/Ralls  | 1            | 16000             | Operator information | 45.42          | Time Based      | 11.53            | 11.73            |
| 6   | MSDBP     | St. Louis     | St. Louis     | 1            | 306647            | Operator information | 567.81         | Time Based      | 337.66           | 858.15           |
| 7   | COLMB     | Columbia      | Boone         | 1            | 123180            | Operator information | 77.98          | Time Based      | 54.81            | 92.63            |
| 8   | MSDFN     | St. Louis     | St. Louis     | 1            | 24174             | Operator information | 25.55          | Time Based      | 14.01            | 35.09            |
| 9   | BROOK     | Brookfield    | Linn          | 1            | 4600              | Operator information | 20.06          | Time Based      | 5.72             | 16.81            |
| 10  | CAPEG     | Cape Girardeau | Cape Girardeau | 1          | 38000             | Operator information | 7.57           | Flow Based      | 2.02             | 1.49             |
| 11  | NEVAD     | Nevada        | Vernon        | 1            | 8000              | Connections with population correction | 41.64          | Time Based      | 16.05            | 45.92            |
| 12  | Anonymous | Facility #1   | -             | 1            | 10559             | Operator information | 7.57           | Time Based      | 3.76             | 3.36             |

\(^a\) Unit: million liter per day (MLD).

\(^b\) Samples were collected during the week of Jan 18th, unit: MLD.

\(^c\) Samples were collected during the week of Jan 25th, unit: MLD.

#### 2.2.2. Plasmid standard preparation

A plasmid carrying a PMMoV gene 180-bp fragment (Table 2) along with a N gene fragment was constructed, purified from *Escherichia coli*, and used as standards for the RT-qPCR assay. The primer pair, COVID19-N 5p and COVID19-N 3p (Table 2), was used to amplify the N ORF fragment from IDT’s 2019-nCoV_N Positive Control plasmid and the N ORF fragments were infused using an InFusion kit (Takara) as described in Robinson et al., 2022. The detail information of plasmid construction and standard curves preparation were described in Text S1.

#### 2.2.3. Quantitative RT-qPCR assay

The TaqMan probe and the primer pairs for N1 and N2 detection from Integrated DNA Technologies (IDT) (Table 2) were chosen based on the CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel (Acceptable Alternative Primer and Probe Sets). Final RT-qPCR one-step mixtures for N1/N2 or PMMoV detection consisted of 5 µL TaqPath 1-step RT-qPCR Master Mix (Thermo Fisher), 500 nM of each primer, 125 nM of the TaqMan probes, 5 µL of wastewater RNA extract, and RNase/DNase-free water to reach a final volume of 20 µL. All RT-qPCR assays were performed in duplicate using a 7500 Fast real-time qPCR System (Applied Biosystems). The reactions were initiated with 1 cycle ofUNG incubation at 25°C for 2 min and then 1 cycle of reverse transcription at 50°C for 15 min, followed by 1 cycle of activation of DNA polymerase at 95°C for 2 min and then 45 cycles of 95°C for 3 s for DNA denaturation and 55°C for 30 s for annealing and extension. The data was collected at the step of 55°C extension. The dilution technique was performed on selected wastewater samples to examine the PCR inhibition, and our results showed no PCR inhibition. In addition, the Puro, a non-infectious retroviral virus with a unique engineered sequence (Puro) (Robinson et al., 2022), was used as the internal standard fortified to the sample.

### Table 2

| No. | Name                  | Sequence                                                                 |
|-----|-----------------------|--------------------------------------------------------------------------|
| 1   | PMMoV gene fragment   | 5’TTTTCCCCGATGTGTA ATACATTTAGGGTAGATCCATTTGCGGAGACAAA GGTAGTGGATCTGTTTTTCAATGAGAGTGGT TTGACCTAAGTTGAGAGCGCTACGGAAGCAA ATGTGGACAGTTTTGCAACGGACAGGAA GAAAGTTGCTTGAAGATTTGAG 3’ |
| 2   | COVID19-N 5p          | 5’ATGTTGTAATGAACGCCCCCAAAATCAGCGG 3’                                    |
| 3   | COVID19-N 3p          | 5’TATGCTTGAAGGATCCCAAGCGCAGCAGCAG 3’                                    |
| 4   | 2019-nCoV_N1-Probe    | 5’ACCCGCGATTACGTGTTGCGAACCC 3’                                         |
| 5   | 2019-nCoV_N1-F        | 5’GACCGAAATCTCGGAAAAT 3’                                                |
| 6   | 2019-nCoV_N1-R        | 5’TCTGGTATTCCGCTGGTATCTGCGTGC 3’                                       |
| 7   | 2019-nCoV_N2-Probe    | 5’FAM 5’ACACATTGGCCCCAGCCCGTTCACG 3’ BHQ1                               |
| 8   | 2019-nCoV_N2-F        | 5’TTCACAAAATTGCCCCCGCCAAGAA 3’                                         |
| 9   | 2019-nCoV_N2-R        | 5’GGGCGACATCCGGAAGAA 3’                                                |
| 10  | PMMoV Probe           | 5’VIC-5’GGTGGTAGATCCGTTGG 3’-QSY                                        |
| 11  | PMMoV Forward         | 5’GGGTGATCCGTTGG 3’                                                    |
| 12  | PMMoV Reverse         | 5’GGAACCTGCTCCTTGGTGAGG 3’                                              |

\(^a\) Acceptable Alternative Primer and Probe Sets: https://www.cdc.gov/coronavirus/2019-ncov/downloads/List-of-Acceptable-Commercial-Primers-Probes.pdf.
wastewater samples to examine if there is RT-PCR inhibition due to the possible PCR inhibitors as quality assurance and quality control. The detail of Qa/Qc were included in Text S2, Table S3 and Figs. S1, 2, 3.

2.3. Quantification of biomarkers

2.3.1. Detection of PMMoV viral concentration

The TaqMan probe (PMMoV Probe) and the primer pair (PMMoV Forward and PMMoV Reverse, Table 2) were designed, purchased from Fisher Scientific (USA) and used to target the PMMoV RNA. The specificity of primers and probe were tested by BLAST analysis (NCBI) to prevent known nonspecific binding targets that could be obtained in a human specimen. The PMMoV concentration in the wastewater sample is determined by the quantitative RT-qPCR assay as described above. The PMMoV limit of detection (LOD) was determined when the concentration that can be detected with 95% certainty (Bustin et al., 2009) (Text S2 and Table S3).

2.3.2. Extraction of 5-hydroxyindoleacetic acid

The wastewater was filtered through a 0.2 µm Whatman® Anotop® filter (Fisher Scientifics, USA). Twenty ml of filtered wastewater was fortified with 20 µL of 100 ppm 5-HIAA-[13C6] followed by solid-phase extraction (SPE) using Waters Oasis HLB SPE cartridge (500 mg) (Waters Milford, MA). The extracts on the SPE cartridge were eluted with the mixture of 50% acetonitrile (ACN) and 50% methanol. The samples were reconstituted with ACN after evaporation. Samples were stored at -20 °C until analyzed by the high-performance liquid chromatography-tandem mass spectrometry (LC-MSMS) analysis.

2.3.3. Extraction of creatinine, caffeine, and paraxanthine

One thousand and six hundred µL of a subsample wastewater was centrifuged for 10 min at 10,000 rpm. One mL of the supernatant was transferred and spiked with 10 µL of formic acid followed by a vortexing vigorously. The mixture was centrifuged at 10,000 rpm for 10 min. Seven hundred fifty µL of supernatant was mixed with 750 µL of LC-MSMS buffer (10 mM ammonium acetate and 0.1% formic acid in water) followed by fortification of 20 µL of 76 ppm caffeine-C13 or creatinine-D3. The mixture was filtered through a 0.2 µm Anotop PTFE filter before the LC-MSMS analysis.

2.3.4. Liquid chromatography-tandem mass spectrometry analysis

All of the analytical standards were purchased from Sigma-Aldrich (St. Louis, MO, USA) except 5-Hydroxyindoleacetic acid-[13C6] (5-HIAA-[13 C6]) (>98%) was purchased from IsoSciences (Ambler, PA, USA). The HPLC grade methanol and acetonitrile used in these experiments were purchased from Sigma-Aldrich (St. Louis, MO, USA). The quantification of 5-HIAA, creatinine, caffeine, and paraxanthine was performed by a Waters Alliance 2695 High Performance Liquid Chromatography (HPLC) system coupled with Waters Acquity TQ triple quadrupole mass spectrometer (MS/MS). The analytes were separated using a Phenomenex (Torrance, CA) Kinetex C18 (100mm x 4.6 mm; 2.6 µm particle size) reverse-phase column. The mobile phase consisted of (A)10 mM ammonium acetate and 0.1% formic acid in water and (B) 100% acetonitrile. The gradient conditions were 0–0.3 min, 2% B; 0.3–7.27 min, 2–80% B; 7.27–7.37 min, 80–98% B; 7.37–9.0 min, 98% B; 9–10 min 98–2% B; 10.0–15.0 min, 2% B at the flow rate of 0.5 mL/ min. The ion source in the MS/MS system was electrospray ionization (EI) operated in either positive or negative ion mode with a capillary voltage of 1.5 kV. The ionization sources were programmed at 150°C and the desolvation temperature was programmed at 450°C. The optimized collision energy, cone voltage, molecular and product ions of biomarkers are summarized in Table 3.

2.4. Normalization of SARS-CoV-2 concentration with biomarker concentration

The CDC/ European Commission recommended PMMoV population normalization scheme has been widely used especially when (1) the metadata (e.g., flow rates, size of population) are not available or not reliable (fluctuated flow rates, population dynamic, dilution from industrial or natural discharges) or (2) samples were collected through grab sampling, manhole, or passive sampling process. The conventional normalization process simply divides the concentration of viral load by the concentrations of PMMoV, and generate a unitless normalized values (Ai et al., 2021; D’Aoust et al., 2021; Feng et al., 2021; Greenwald et al., 2021).

Our approaches are similar to the currently widely used population normalization scheme (unitless), but they are more advanced by normalizing the viral concentrations into copies/person, so it could be directly compared against the infection rates (clinical cases/population) within each sewershed. In addition, since the outcome is a universal unit – copies/person, it will allow us to compare the results among different WWTF facilities. Our approaches could be particularly useful when the metadata (e.g., flow rates, size of population) are not available or not reliable under the same scenarios described above.

Two approaches were proposed to normalize SARS-CoV-2 concentration in the wastewater using the concentrations of biomarkers and established regression functions from the linear regression models. The direct approach utilized the ratio of population contributing to the wastewater to wastewater volume (P, size of the population/L) estimated by biomarker concentration to directly normalize SARS-CoV-2 load, which is similar to the PMMoV normalization scheme recommended by CDC as well as European Commission and adopted by the majority of the wastewater-based epidemiology programs. As for indirect approach, the nominator of biomarker concentration was utilized and replaced with the estimated population during the normalization process. Both approaches assumed that the biomarker load is proportional to the population in the wastewater composite (Fig. 1). This section presented the methods of (1) determining the regression functions and (2) normalizing SARS-CoV-2 concentrations using biomarker concentrations are presented.

2.4.1. Relationships between biomarker concentration and ratio of population contributing to the wastewater to its volume

The ratio of population contributing to the wastewater and wastewater volume (P, size of the population/L) is expressed as

| No. | compound | RT (min) | ES | MS1 (m/z) | MS2 (m/z) | Cone Voltage (V) | Collision Energy (eV) |
|-----|----------|---------|----|-----------|-----------|-----------------|----------------------|
| 1   | Caffeine | 6.273   | ES+ | 195.05    | 138.12    | 45              | 22                   |
| 2   | Caffeine-13C6 | 6.167 | ES+ | 198.04    | 140.07    | 45              | 22                   |
| 3   | Paraxanthine | 5.715 | ES+ | 181.06    | 124.11    | 45              | 22                   |
| 4   | 1,7-Dimethylxanthine-(dimethyl-D3) | 5.72 | ES+ | 187       | 127.1     | 30              | Tune                 |
| 5   | 5-hydroxyindoleacetic acid | 6.135 | ES+ | 192       | 146       | 30              | 14                   |
| 6   | 5-hydroxyindoleacetic acid-13C6 | 6.145 | ES+ | 198       | 152       | 30              | 14                   |
| 7   | Creatinine | 2.189 | ES+ | 114.05    | 44.06     | 30              | 14                   |
| 8   | Creatinine-D3 | 2.189 | ES+ | 117       | 47        | 30              | 14                   |
A. Normalization with Metadata

| Information available |
|------------------------|
| 1. Daily flow volume (10^6 L) |
| 2. [SARS-CoV-2] (copies 10^6 L) |
| 3. Population |

B. Normalization with biomarkers

| Information available |
|------------------------|
| 1. Biomarker (mg 10^6 L) |
| 2. [SARS-CoV-2] (copies 10^6 L) |

Fig. 1. Normalization processes of determining SARS-CoV-2 load per capita. (A) When the population size, daily flow volume and viral concentration of the metadata are used in the normalization process. (B) When the real-time population size of the sewershed is estimated using regression functions developed from the correlation between biomarker and population size from metadata in direct or indirect approach.

\[
(P_i) = \frac{P_i}{V_i}
\]  \hspace{1cm} (1)

in which, \(P_i\) and the daily flow volume \(V_i\) (Litter) for WWTP j are provided in metadata (Table 1). The \(P_i\) is modeled as

\[
[B_i] = \beta_i \times V_i + \varepsilon_i
\]  \hspace{1cm} (2)

where \([B_i]\) is the concentration of biomarker \(i\) in WWTP j sample, the corresponding \(P_i\), the error term \(\varepsilon_i\) and the estimated parameter \(\beta_i\) for biomarker \(i\). The error term accounts for differences in biomarker concentration from daily variations at the locations. To avoid any skewness, Log-transformed population and biomarker concentrations were further used to fit a linear regression model. The Pearson’s correlation coefficient \(r\) was calculated.

2.4.2. Relationships between biomarker loads and population size

Daily flow volume was taken into consideration before the relationship between daily biomarker load and the population contributing to the wastewater was examined. The biomarker load of biomarker \(i\) for WWTP \(j\), \(B_i\), was calculated as

\[
B_i = \frac{[B_i]}{V_i}
\]  \hspace{1cm} (3)

in which, \([B_i]\), the biomarker \(i\) concentration in WWTP \(j\) wastewater samples, was determined by LC-MSMS. The population \(P\) is modeled as

\[
B_i = \beta \frac{P_i}{V_i} + \varepsilon_i
\]  \hspace{1cm} (4)

where \(B_i\) is the daily \(i\) biomarker load, \(P_i\) the population from metadata at WWTP \(j\).

2.4.3. Developing the normalization scheme derived from metadata

According to the CDC’s guideline, the normalization of SARS-CoV-2 load (copies/person/day) is expressed and calculated as

\[
\frac{\text{Viral load}}{\text{Population}} = \frac{[N1,N2]_{\text{SARS}} \times E \times (V \times 10^6)}{P}
\]  \hspace{1cm} (5)

in which, \([N1,N2]_{\text{SARS}}\) is the average of replicated N1 and N2 concentrations \((n=4)\) in the wastewater samples. \(E\), concentration factor, 350, transforms unit of concentration from copies/μL of RNA to copies/L of wastewater. Daily flow volume \(V\) (Litter) and population \(P\) are provided in Metadata. In the last line, all variables and constants are designated as normalization coefficient 0 \((C_0)\) except \([N1,N2]_{\text{SARS}}\). The unit of normalized SARS-CoV-2 load per capita turns into copies per person.

2.4.4. Developing the normalization scheme derived from the relationship between biomarker concentration and \(P\)

\(P_i\), the ratio of population contributing to wastewater and wastewater volume, estimated by biomarker concentration in the wastewater was utilized in the direct normalization approach. Since \(P_i\) is expressed as \(\frac{P_i}{V_i}\) in Eq. (1), the correlation between the \([B_i]\) biomarker \(i\) concentrations and \(P_i\) is expressed as

\[
\frac{1}{B_i} \sim \frac{1}{P_i} \sim \frac{V_i}{P_i}
\]  \hspace{1cm} (6)

in which population \(P_i\) and daily flow volume \(V_i\) were estimated using biomarker \(i\) concentration in the Eq. (2). The reciprocal of the estimated population \(P_i\) and daily flow volume \(V_i\) derived from \(P_i\) were unitized in SARS-CoV-2 load normalization process:

\[
\frac{\text{Viral load}}{\text{Population}} = \frac{[N1,N2]_{\text{SARS}} \times E \times (V_i' \times 10^6)}{P_i'}
\]  \hspace{1cm} (7)

\[ \text{Copies per capita}_S = \frac{[N1,N2]_{\text{SARS}} \times E \times (V_i' \times 10^6)}{P_i'} \]
in which, the \( P \) and \( V \) in line 2 are replaced with \( P' \) and \( V' \) in Eq. (6) resulting in line 3. Since \( \frac{P}{V} \) correlated with \( [B_i] \), \( \frac{V'}{P'} \) is replaced with \( [B] \) in line 4. The \( (P') \) is further modeled using \( [B] \) in line 5. Except \( [N1, N2]_SARS \) all of variables and constants were designated as normalization coefficient 1, \( C_{1(i)} \), for biomarker \( i \) in the direct approach. The \( C_{1(i)} \) was further standardized by \( C_0 \) as

\[
\text{Fold change} = \frac{C_{1(i)}}{C_0} \tag{8}
\]

The fold change was utilized to assess the fitness, precision, and the variability of the biomarkers.

2.4.5. Developing the normalization scheme derived from the relationship between biomarker loads and population

The population estimated by biomarker loads in the wastewater were used in the indirect biomarker to fall into the linear range of the correlation:

\[
[B_i] \times 10^9 \text{ g/L} \tag{9}
\]

\[
= \frac{[B_i]}{10^9} \text{ g/L}
\]

in which, \( [B_i] \) is the concentration of biomarker \( i \) (µg/L or copies/L). The population was estimated using \( [B] \times 10^9 \) as \( B \) in the Eq. (4), and the unit of estimated population contributing to wastewater (\( (P') \)) became person/L. The \( (P') \) estimated by biomarker \( i \) is further utilized in SARS-CoV-2 load normalization below:

\[
\text{Viral load} = \frac{[N1, N2]_SARS \times E \times (V' \times 10^9)}{P'} \tag{10}
\]

\[
= \frac{[N1, N2]_SARS \times E \times (V' \times 10^9)}{[B] \times 10^9 \times (V' \times 10^9)}
\]

\[
= \frac{[N1, N2]_SARS \times E \times (V' \times 10^9) \times 10^9}{[B] \times 10^9 \times (V' \times 10^9)}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[B] \times 10^9}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[P']}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[P']}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[P']}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[P']}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[P']}
\]

\[
= \frac{[N1, N2]_SARS \times E \times 10^9}{[P']}
\]

in which, the daily flow volume and constants in both numerator and denominator were canceled out in line 3, which resulted in line 4. Except \( [N1, N2]_SARS \), all of variables and constants were designated as normalization coefficient 2, \( C_{2(i)} \), for biomarker \( i \) in the indirect approach. The \( C_{2(i)} \) was further standardized by \( C_0 \) as

\[
\text{Fold change} = \frac{C_{2(i)}}{C_0} \tag{11}
\]

2.5. Validation of normalization coefficients

The regression function of two approaches were established to normalize SARS-CoV-2 load using the 24 samples collected in January 2021 (Table 1). Samples collected from 64 WWTPs in May 2021 (Table S1) were utilized as testing data set to validate the estimation of the normalization coefficients (\( C_{1(i)} \) and \( C_{2(i)} \)) from two approaches. During the validation, \( C_0 \) was calculated using Metadata in Eq. (5). The \( C_{1(i)} \) and \( C_{2(i)} \) were calculated using the concentration of CAF and PARA with Eqs. (7) and (10), respectively, followed by standardization with \( C_0 \) to evaluate the fitness, precision, and the variability.

2.6. Estimation of population contributing to the wastewater

2.6.1. Linear regression model

To determine the accuracy and precision of population estimated by different biomarkers, the log-transformed biomarkers loads (\( n \), 24), collected from 12 WWTPs across the State of Missouri (Table 1), were used as predictor variable to fit the linear regression model in R.

\[
P = \beta_0 + \epsilon_i \tag{12}
\]

Nineteen of the data points (approximately 80%) was randomly selected as training data set to fit the model, and the rest 5 data points were used as test data set. The adjusted \( R^2 \) and the mean square error (MSE) were utilized to evaluate the model fitting and prediction accuracy, respectively. A k-fold cross-validation (\( k = 5 \)) was performed to eliminate the poor prediction from the outliers and determine the overall predictive capability of the model based the 5-fold cross-validation MSE (Refaeilzadeh et al., 2009).

2.6.2. Estimation of real-time populations for City of Columbia (college Town) and a Tourist Town

The population contributing to the sewershed was expected to fluctuate over the surveillance period due to tourism, weekday commuters, temporary workers, and quarantine etc. To monitor the population fluctuation, wastewater samples were collected from the WWTPs of City of Columbia (college town) and a tourist town over 10 time points (Table S2). The PARA load at each given time was calculated using PARA concentration and the daily flow volume reported in the metadata as in Eq. (3). The population at each given time was further estimated using the linear regression model built from Eq. (4) and the calculated PARA loads.

2.7. Relationships between SARS-CoV-2 load in wastewater and clinical prevalence

The weekly average of SARS-CoV-2 clinical case numbers in City of Columbia was collected by the Missouri Department of Health and Senior Services from May to September 2021. \( C_p \) was calculated using metadata in Eq. (5); \( C_{2(i)\text{PARA}} \) was calculated using the concentration of PARA in Eq. (10). SARS-CoV-2 concentration was normalized by \( C_p \) and \( C_{2(i)\text{PARA}} \) depending on the scenarios: (1) SARS-CoV-2 load per capita normalized by metadata versus clinical cases normalized by metadata, (2) SARS-CoV-2 load per capita normalized by \( C_{2(i)\text{PARA}} \) versus clinical cases normalized by metadata and (3) SARS-CoV-2 load per capita normalized by \( C_{2(i)\text{PARA}} \) versus clinical cases normalized by PARA-estimated population using Eq. (12). Spearman’s correlation analysis was performed to examine the correlation between normalized SARS-CoV-2 concentration and one-week average clinical case numbers.
3. Results

3.1. Relationships between biomarkers and population

Twenty-four samples collected from 12 WWTPs in the state of Missouri (Table 1) were used to assess the correlation between biomarkers and Pr (ratio of population contributing to wastewater to wastewater volume) using Eq. (2) or biomarker and population using Eq. (4). The linear regression models were fitted by either the biomarker concentration and Pr in Eq. (2) or biomarker loads and population in Eq. (4). The R square ($R^2$) represents the variation of population/Pr explained by the model. The Pearson’s correlation coefficient ($r$) represents the strength of the correlation.

The concentrations of CAF showed the highest correlation (Pearson coefficients, $r = 0.810$) with the Pr, followed by the concentrations of PARA ($r = 0.774$), PMMoV ($r = 0.598$), 5-HIAA ($r = 0.59$), and CRE ($r = 0.06$) (Fig. S4 and Table S4). Log-transformation has been widely used to process the skewed data. It helps to decrease the variability of data and make data conform more closely to the normal distribution (Feng et al., 2013). After log-transformation, the $r$ was increased to 0.886 for CAF, 0.861 for PARA, 0.720 for 5-HIAA, and 0.707 for PMMoV (Fig. 2), however, it was not improved for CRE.

The daily load of CAF exhibited the highest correlation ($r = 0.99$) with population, followed by the daily load of 5-HIAA ($r = 0.98$), PMMoV ($r = 0.98$), PARA ($r = 0.97$), and CRE ($r = 0.22$) (Fig. S5 and Table S5). Similarly, log-transformation significantly improved the correlation of all five coefficients. The PARA and CAF daily load showed the highest correlation ($r = 0.97$ and 0.97, respectively) with the population, followed PMMoV load ($r = 0.92$), 5-HIAA load ($r = 0.87$), and CRE load ($r = 0.33$) after log-transformation (Fig. 3).

3.2. Comparison of normalization coefficients among different biomarkers

The normalization coefficient ($C_{1(i)}$ or $C_{2(i)}$) calculated from biomarker concentration were utilized to normalize SARS-CoV-2 viral load. A reliable biomarker for population normalization should achieve high precision and low variability, meaning that the normalization

![Fig. 2. Log-transformed Pr (ratio of population contributing to wastewater to its volume, population/L) versus biomarker concentration (μg/L) in the wastewater. (A) CAF: caffeine, (B) PARA: paraxanthine, (C) 5-HIAA: 5-hydroxyindoleacetic acid, (D) PMMoV: Pepper Mild Mottle Virus (E) CRE: creatinine. The concentrations of caffeine, paraxanthine, 5-hydroxyindoleacetic acid, and creatinine in 24 wastewater samples (Table 1) were determined by LC-MS/MS analysis and the Pepper Mild Mottle Virus concentration was determined by RT-qPCR as described in Methods and Materials. The population concentrations were calculated using the daily flow volume and population size in Eq. (1). The trendline (dashed line) was calculated using linear regression; $R^2$ represented the percentage of the population concentration variation that is explained by the linear model.](image-url)
coefficient ($C_{1(i)}$ or $C_{2(i)}$ for biomarker $i$) should be comparable to $C_0$ calculated from the population and daily flow volume derived from metadata. Hence, when the normalization coefficients from different biomarkers were standardized by $C_0$ as fold change ($C_{1(i)}/C_0$), the closer to 1 ($y=1$) the fold change is, the higher precision and lower variability the biomarker obtains.

In the direct normalization approach, $C_{1(i)}$ were calculated using the Eq. (7) and biomarker concentrations. CAF outperformed other biomarkers resulting from the lower variation, and higher precision in comparison of the $C_{1(i)}$ of all other biomarkers (Fig. 4 and Table S6). Most of $C_{1(5-HIAA)}$ and $C_{1(PMMoV)}$ among wastewater facilities showed variation above the baseline ($y = 1$), which could result in over-normalization of SARS-CoV-2. The relatively high variation of $C_{1(5-HIAA)}$ and $C_{1(PMMoV)}$ could over-normalize or under-normalize. The $C_{1(CRE)}$ results were not included in this comparison due to its poor correlation with population. Therefore, the results suggested that the CAF should be the most suitable biomarker for the direct normalization approach, followed by PARA, 5-HIAA and then PMMoV at last.

In the indirect normalization approach, the normalization coefficients ($C_{2(i)}$) were calculated with the data-transformed biomarker concentrations in Eq. (9), followed by standardization by $C_0$ and expressed as fold change ($C_{2(i)/C_0}$). The fold change ($C_{2(PARA)/C_0}$) of PARA outperformed other biomarkers due to its lower variation, and higher precision (Fig. 5 and Table S7). Among all biomarkers, CRE exhibited the highest variation and lowest precision. Thus, the most suitable biomarker for the indirect normalization approach would be PARA, followed by CAF, PMMoV and 5-HIAA.

### 3.3. Normalization of SARS-CoV-2 load per capita

The SARS-CoV-2 loads normalized by biomarkers (copies/person) were directly calculated by multiplying the viral concentrations with the normalization coefficient of the corresponding biomarker. Fig. 6 demonstrated the biomarker-normalized viral per capita of each selected facility in the State of Missouri for the week of January 19th and week of January 23rd, 2021. Unlike most of the current SARS-CoV-2 wastewater-based epidemiology programs, our normalization strategy allows us to compare the viral load per capita among the sewersheds cross the geographic gradient. The vulnerability meant high risk of the exposure and transmissibility among the communities due to the

![Graphs showing the relationships between log-transformed population and biomarker load in wastewater](image-url)
potential of the high infection rates (e.g., viral load/per capital in wastewater = high clinical cases/population = high risk). The vulnerable community was identified by comparing the viral load per capita among the sewersheds as function of time. The viral load per capita was determined by the SARS-CoV-2 load normalized by the population, which was estimated with the population biomarkers. For example, viral load per capita at BROOK WWTP in week of January 19th (Fig. 6B and D) was significantly higher than other sewersheds, and the viral load per capita was determined by the SARS-CoV-2 load normalized by the population, which was estimated with the population biomarkers. For example, viral load per capita at BROOK WWTP in week of January 19th (Fig. 6B and D) was significantly higher than other sewersheds, and the viral load per capita was...
Water Research 223 (2022) 118985

S.-Y. Hsu et al.

capita in week of January 19th was much higher than a week earlier (Fig. 6A and C), suggesting that the BROOK was a vulnerable community, that might soon encounter high COVID-19 illness and high infection rate.

3.4. Validation of normalization coefficients

Based on the value of fold change, CAF and PARA achieved the lowest variability and highest accuracy, and precision (Figs. 4 and 5). These normalization approaches were further validated using the wastewater samples collected from 64 WWTPs in the State of Missouri in May 2021 (Table S1). The normalization coefficients, \( C_1 \) (CAF), \( C_1 \) (PARA), \( C_2 \) (CAF) and \( C_2 \) (PARA), for each WWTP was calculated using the established regression functions between CAF/PARA and population (Tables S4 and S5) without metadata. These coefficients were normalized by \( C_0 \) derived from metadata to assess the fitness, precision, and variability.

There was no significant difference between the normalization coefficients of CAF and PARA when the direct approach or indirect approach was applied (Fig. 7, Tables S8 and S9). The mean of fold change of CAF and PARA from direct and indirect approach were both close to 1 (high precision) with few outliers (low variability). These results not only concurred with the results shown in Figs. 4 and 5 but also indicated that the regression functions developed in this study could be used for normalizing SARS-CoV-2 load without metadata in the future.

Fig. 6. Identification of the vulnerable communities using normalized SARS-CoV-2 viral load by biomarkers with either direct or indirect approaches at WWTPs. The direct normalization approach was applied to 12 samples collected in the week of (A) January 19th and (B) January 23rd. The indirect approach was applied to 12 samples collected in the week of (C) January 19th and (D) January 23rd. (Grey: Metadata, yellow: CAF, blue: PARA, green: PMMoV, orange: 5-HIAA; error bars showed standard deviation, \( n = 4 \)). The SARS-CoV-2 load per capita was normalized using the average of duplicated N1 and N2 concentrations at each WWTP and the normalization coefficients of each biomarker in Eq. (7) for direction approach in (A) and (B), or in Eq. (10) for indirect approach in (C) and (D). The viral loads were normalized using metadata in Eq. (5) and included in all graphs for comparison. The data of CRE was not shown due to its poor correlation with population.

Fig. 7. Validation of normalization approaches. The direct approach for (A) CAF and (B) PARA and the indirect approach for (C) CAF and (D) PARA were applied and shown for validation. In the box plots, the upper whisker represents the maximum, the lower whisker the minimum; "X" represents the mean and open circles are the outliers. The PARA and CAF concentrations in 64 wastewater samples collected from WWTPs in the State of Missouri (Table S1) were quantified by LC-MS/MS, and the normalization coefficients, \( C_0 \), \( C_{1(i)} \) and \( C_{2(i)} \), were calculated as described in Methods and Materials. The fold changes \( (C_{1(i)} / C_0 \) or \( C_{2(i)} / C_0 \) were used to standardize \( C_{1(i)} \) and \( C_{2(i)} \).
3.5. Estimation of real-time population contributing to the wastewater

The precision of real-time biomarker-estimated populations were assessed by fitting regression models with the biomarker loads using R program. PARA achieved the highest adjusted R square, followed by CAF, 5-HIAA, PMMoV and CRE (Table 4). The PARA showed the lowest mean square error (MSE), which is the parameter used for assessing the prediction accuracy by the developed model and it was increased in the order of CAF, PMMoV, 5-HIAA and CRE. Again, PARA obtained the lowest 5-fold cross-validation MSE, suggesting that PARA is the most suitable biomarker for estimating the population.

To accurately normalize SARS-CoV-2 levels per capita over time, the populations at a college town and a tourist town were estimated using the PARA concentrations in wastewater samples collected through May to early September in 2021. When the daily flow volume was available, the real-time population was predicted by the biomarker loads using the established biomarker loads vs. populations regression functions in Eq. (3) (Table S5). The results showed the real-time population dynamic of population at City of Columbia, especially in late May, August, and early September (Fig. 8A). The variation of estimated populations in Columbia were from -36% to 8% compared to the population reported in Metadata. The change in the real-time population from May to early September in a tourist town were observed in similar pattern (Fig. 8B).

3.6. Correlation between SARS-CoV-2 load per capita and clinical prevalence

It was demonstrated in Fig. 9 that the relation between SARS-CoV-2 levels in the wastewater and clinical cases could be misrepresented without a proper normalization using a reliable population marker. This is mainly attributed to that the population in the City of Columbia was constantly fluctuating over the surveillance period (Fig. 8A). The Spearman’s rank correlation was performed to understand the correlation between viral loads and prevalence data (Udovičič et al., 2007). Spearman’s correlation coefficient, \( \rho \), represents strength of the correlation between viral loads and prevalence data.

For instance, the correlation between the average weekly case number and the SARS-CoV-2 concentration over time was insignificant (\( \rho = 0.5152, p < 0.1 \)) before normalization (Fig. 9A). The \( \rho \) was reduced to 0.47 (\( p < 0.1 \)) after the viral concentration and clinical case number were both normalized by the fixed population from the metadata (through population census) (Fig. 9B). Similarly, as the viral concentration normalized by PARA-estimated population (copies/person) plotted against the clinical case numbers normalized by metadata (case number/100, 000 people), \( \rho \) dropped to 0.50 (\( p < 0.1 \)) (Fig. 9C). In contrast, when both viral load and clinical case number were properly normalized by PARA-estimated population, the correlation was positive and moderate (\( \rho = 0.59, p < 0.05 \)) (Fig. 9D).

4. Discussion

4.1. Population Biomarker selection

Our findings suggested the chemical marker, PARA, is more reliable population biomarker than PMMoV, due to its (1) better population indicators with higher accuracy, lower variability (Figs. 4 and 5) and higher temporal consistency (Fig. 8), (2) very limited exogenous sources (Summers et al., 2015), (3) higher extraction efficiency with low variability (Table 5), (4) higher stability in wastewater (Choi et al., 2020), (5) resistant to chemicals in the wastewater (Armanious et al., 2016), and (6) low sample volume requirement with simple sample preparation process (Table 5) although PMMoV is recommended by US CDC and European Commission as population fecal biomarker to normalize SARS-CoV-2 concentrations.

4.1.1. Population indicators

Pepper Mild Mottle virus (PMMoV) and caffeine are commonly found in human diet. The log-transformed PARA, caffeine metabolite, daily load in wastewater demonstrated better correlation with population (\( r = 0.97 \)) as compared to PMMoV (\( r = 0.92, \) Fig. 3), resulting PARA outperformed PMMoV no matter which normalization was applied. The variability in biomarker shedding could result from different social and cultural variations influencing food habits. The per capita availability of fresh bell pepper and processed Chile pepper were 11.3 and 7.9 pounds in 2021 in the United States (Davis et al., 2022). An early study (Holezer et al., 2012) reported that women and members of high-income households are the most likely subpopulations to consume peppers. Regarding ethnicity, Hispanic is more likely to consume peppers as compared to white, black and others. With respect to age, elders >50 years old are more likely to consume peppers as compared to adults, aged 14 to 49 years, and children under 14 are less likely to do so as compared to adults.

Although the metabolite of the caffeine, PARA, has been recognized as the reliable population biomarker, caffeine intake could be affected by the ethnicity and age. For example, non-Hispanic black individuals

---

**Table 4**

Estimation of population using biomarker loads.

| Biomarkers | p value | Adjusted \( R^2 \) | \( \text{MSE} \) | k-fold Cross-validation MSE |
|------------|---------|------------------|----------------|------------------|
| CAF        | 0.00    | 0.938            | 0.0723         | 0.0251           |
| PARA       | 0.00    | 0.9404           | 0.0516         | 0.0182           |
| 5-HIAA     | 0.00    | 0.8531           | 0.6124         | 0.1065           |
| PMMoV      | 0.00    | 0.9043           | 0.5125         | 0.0501           |
| CRE        | 0.10    | 0.1189           | 0.9400         | 0.2517           |

\( b \) MSE: mean square error.

\( a \) The biomarker loads, and population were transformed using log10.

\( c \) k-fold Cross-validation was performed when \( k = 5 \) and averaged MSE was calculated.

---

**Fig. 8.** Estimation of real-time population in the college town and the tourist town. (A) College town (B) Tourist town (blue triangle: population estimated using PARA, orange circle: population reported by Metadata). The PARA concentrations in 10 wastewater samples collected from WWTWs in City of Columbia and a tourist town (Table S2) were quantified by LC-MS/MS as described in Methods and Materials and further converted to daily PARA load using daily flow volume from metadata. The population was estimated using the daily PARA load using the developed regression function (Table S4).

---

**Fig. 9.** Correlation between SARS-CoV-2 load per capita and clinical prevalence (copies/PARAs) using PARA load (copies/PARAs) and clinical case number (per 100,000 people) for (A) College town (Fig. 9A) and (B) Tourist town (Fig. 9B). Spearman’s rank correlation coefficient \( \rho \) represents strength of the correlation between viral loads and prevalence data.
consumed the smallest amounts (80±2 mg/d), non-Hispanic white individuals consumed the greatest amounts (194±3 mg/d), and Asian individuals (126±7 mg/d) and Hispanic individuals consumed intermediate amounts (127±3 mg/d). Middle-aged individuals (aged 50 to 54 years) consumed more caffeine (211±6 mg/d) than younger (107±4 mg/d, aged 20 to 24 years) and older individuals (153±4 mg/d, aged 75 to 79 years) (Lieberman et al., 2019). Daily caffeine intake from beverages was 14 mg/day in children aged 1 to 5 years old, and 22 mg/day in children aged 6 to 9 years old (Knight et al., 2004). Interestingly, the caffeine intake was not associated with physical activity, economic

Fig. 9. The correlation between normalized SARS-CoV-2 loads in wastewater and the clinical reported case numbers. (Orange dashed line: clinical case, blue solid bar: normalized N1/N2 average concentration/load). The PARA concentrations in 10 wastewater samples collected from WWTP in City of Columbia (Table S2) were quantified by LC-MS/MS as described in Methods and Materials and applied in Eq. (10) to normalize viral load using indirect approach. (A) Viral concentrations and clinical cases before normalization (B) Both viral load per capita and clinical cases normalized using metadata. (C) Viral load per capita normalized by PARA load and clinical cases normalized by Metadata (D) Both viral load per capita and clinical cases normalized by PARA loads. The Spearman’s correlation was performed to examine the correlation between normalized SARS-CoV-2 and clinical case numbers; \( \rho \) represented the strength of the correlation.

Table 5
Comparison of selected biomarkers in this study.

|                | CAF             | PARA           | 5-HIAA          | PMMoV           | CRE              |
|----------------|-----------------|----------------|-----------------|-----------------|-----------------|
| **Stability in Wastewater** | Stable (Choi et al., 2020; O’Brien et al., 2017) | Stable (Choi et al., 2020) | Stable (Thai et al., 2019) | Poor (Thai et al., 2014) | Poor (Thai et al., 2014) |
| **Storage stability** | Stable > 40 days | Stable > 40 days | - | Poor (half-life 6-10 days) | - |
| **Recovery/Extraction Rate** | \( ^1 \) 101±7% | \( ^9 \) 92±3% | \( ^78 \) ± 19% | 10%-45%± 40%-50% | \( ^2 \) 123±31% |
| **Instrumental LOD** | \( ^3 \) 1.06 \( \mu \)g/L | \( ^3 \) 0.72 \( \mu \)g/L | \( ^4 \) 14.74 \( \mu \)g/L | 20 copies/\( \mu \)L | \( ^5 \) 1.19 \( \mu \)g/L |
| **Signal inhibition** | No | No | No | Sensitive | No |
| **Concentration in wastewater** | 47.3±22.9 \( \mu \)g/L | 4.2±2.5 \( \mu \)g/L | 13.5±5.5 \( \mu \)g/L | 77.9834 copies/\( \mu \)L | 102.8±120.4 \( \mu \)g/L |
| **Sample Volume** | 1.5-2 mL | 1.5-2 mL | 25-50 mL | 50 mL | 1.5-2 mL |
| **Sample Preparation time** (for 12 samples) | 30 min | 30 min | 2-3 h | 3-4 h | 30 min |
| **Analysis time** | 15 min per sample | 15 min per sample | 15 min per sample | 2 h for 64 samples | 15 min per sample |
| **Other exogenous sources** | Disposal of the coffee or caffeinated products (Summers et al., 2015) | Microbial degradation of caffeine (small amount) | - | Ground water, agriculture soils, fertilizers | - |

\( ^{a} \) The recovery rate was calculated from the isotope fortified in wastewater samples.

\( ^{b} \) The limit of detection of LC-MS/MS method as described in the Material and Method.

\( ^{c} \) The limit of detection of RT-qPCR assay as described in the Supplemental information.
status, education level, or employment status (Lieberman et al., 2019).

4.1.2. Exogenous sources

Unlike CAF and PMMoV, PARA is the metabolite product generated through the human consumption of the caffeinated products (coffee, tea and caffeinated drinks), indicating that human is the major source contributing PARA in the wastewater. In humans, 80% of caffeine is metabolized into paraxanthine (Martinez Bueno et al., 2011). The production of the PARA could be also attributed to the microbial degradation of caffeine in the environments, however since it is not the predominant microbial degradation pathway, the amount of PARA produced through this process is very limited (Summers et al., 2015). Therefore, we could assume that PARA loading in WWTP was mostly generated through human consumption of caffeine.

The exogenous sources of PMMoV could be from chicken and seagull feces (Rosario et al., 2009). The PMMoV widely detected in the groundwater, irrigation water and surface water (rivers, ponds) (Asami et al., 2016; Hamza et al., 2011; Kuroda et al., 2015; Rosiles-Gonzalez et al., 2017) also suggested that PMMoV might have other exogenous sources that could contribute to wastewater. In addition, recently, several SARS-CoV-2 wastewater surveillance projects in the U.S. have reported the increased levels of PMMoV after the major stormwater events. Further investigation suggested the potential exogenous sources of the PMMoV from agricultural soils, suspended sediments and fertilizers (personal communication).

4.1.3. Extraction efficiency

variation in the extraction rates (Ahmed et al., 2020b). Variations in the extraction rates of PMMoV that have been widely reported is another drawback (Feng et al., 2021; Kato et al., 2018; LaTurner et al., 2021). Feng et al. reported a recovery of 45±26% PMMoV using direct extraction with HA filters. The PMMoV was also poorly correlated with the recovery of the SARS-CoV-2 enveloped virus. Similarly, Kato et al. reported a wide variability of the PMMoV recovery efficiencies with typical recovery rates only greater than >10% when concentrating using electronegative filters (Kato et al., 2018). The high variability among different concentration techniques for PMMoV analysis, including direct extraction, HA filtration, filtration with bead bearing, PEG precipitation, and ultrafiltration have been illustrated by LaTurner et al. (2021). The coefficient of variation (%CV) for these concentration techniques range from 25.9% to 49.8%. Feng et al. reported that the variability in the PMMoV extraction rates might have contributed to the decreased correlation coefficient between the normalized SARS-CoV-2 concentration and the clinic cases in most of WWTP facilities reported by previous studies (Feng et al., 2021). Among the genic fecal markers, although PMMoV has demonstrated a less variable RNA signal compared to Bacteroides 16S rRNA or human eukaryotic 18S rRNA, the variability of PMMoV has demonstrated a less variable RNA signal compared to PMMPoV (Table 5). The average recovery rates of CAF and PARA in our study were 101% and 92% with standard deviation of ±7% and ±3%, respectively, similar to 73% to 109% for CAF and its metabolites reported by Driver et al. (2020). Both CAF and PARA were found to be relatively stable in the sewer system (Choi et al., 2020). The CAF and PARA have several unique characteristics that are critical to serve as the reliable chemical fecal population markers. They are highly soluble in water (13 g L⁻¹) with a very low hydrophobicity (octanol-water coefficient log Kow = −0.07), insignificant volatility and its half-life is about 10 years (Buerge et al., 2006; Chen et al., 2012; Froehner and Martins, 2008; Lin et al., 2009). Due to the high polarity and water solubility, CAF and PARA will less likely to adhere to the solids fraction of wastewaters via electrostatic and/or hydrophobic partitioning effects as the PMMoV biomarker described by Armanious et al. (2016). As the wastewater stored at -20°C, the PARA could be stable for at least 4 weeks or more (Choi et al., 2020; Senta et al., 2015). With our new modified direct methanol dilution extraction protocol (50% methanol), we anticipate that the CAF and PARA extracts could be stable beyond several months when they are stored at -20°C under the 50% methanol sterilized solution (Buerge et al., 2003).

4.1.6. Sample preparation and analysis

The sample volume required for analysis for PARA is less than 2 mL (0. 1ml with a modified methanol extraction protocol), that is significantly less than 25-50 mL sample volume required for PMMoV analysis (Table 5). Another advantage for using PARA as the fecal marker is that it required less sample preparation time and processes. An average sample preparation time for PARA analysis was less than 30 min/6 samples, with new modified methanol extraction protocols, it could be further reduced to 10 min/6 samples, while the sample preparation time (e.g., extraction and concentration) for PMMoV analysis often takes approximate 3 h.

PMMoV can be used simultaneously quantified as the targets SARS-CoV-2 viral nucleic acid using the RT-qPCR multiplex platforms, while CAF and PARA were quantified using LC-MS/MS separately. Despite that no inhibition observed in the one step RT-qPCR assay in our study, RT-qPCR inhibition have been reported by several studies (Kato et al., 2018). Quality control internal standards, and dilution protocols are often required to account for any PCR inhibition. Incorporation of the internal positive control, such as a modified targeted gene sequence or CGMMV are often required to correct the variation in the extraction efficiency plus any potential inhibition (Kato et al., 2018). Other emerging PCR-based techniques, such as RT-droplet digital PCR (RT-ddPCR) and digital qPCR, were also used for quantifying PMMoV (Hinkle et al., 2022, p.). Nevertheless, they are less affordable, and their implementation in the WBE pipeline requires more time.

Without LC-MS/MS, CAF and PARA could be quantified by other less-expensive alternative analytical techniques, such as gas chromatography–mass spectrometer (GC-MS), high-performance liquid chromatography coupled with photodiode-array detector (HPLC-PDA) due to their much higher concentrations in the wastewater sample (Chen et al., 2002; Thomas and Foster, 2004). Additionally, if the population within sewershed did not fluctuate over time, quantification of PARA could be performed at core facilities equipped with LC-MS/MS once a while.

4.1.7. Other biomarkers

Other biomarkers selected in this study did not meet the criteria of population biomarker. The levels of the creatinine, the metabolite of muscle, did not correlate with population, consistent with the results reported by Thai et al. (2019, 2014). The poor correlation could be due to its instability in wastewater treatment designs and processes, high variance of intra- and extra-individual excretion (Daughton, 2012; Thai et al., 2014). The 5-HIAA, one of the major metabolites of serotonin, correlated with population well and it has been reported to be stable in wastewater (Thai et al., 2019). Nevertheless, the low concentrations in
the wastewater and the observed coeluted interferences in the LC/MSMS analysis, the time required for sample preparation and cleanup, particular the time-consuming concentration and cleanup processes through solid-phase extraction (SPE), make the 5-HIAA not an ideal marker candidate for real-time and rapid analysis. In addition, a sensitive tandem mass spectrometer is the only option for quantifying the 5-HIAA in the wastewater due to its low sub-ppb to ppb concentration range, while Although CAF and PARA shared similar characteristics in wastewater, CAF loading might result from discarded caffeinated products, and therefore, make PARA more desirable population biomarker.

4.2. Normalization of SARS-CoV-2 load and method validation

The utility of chemical biomarkers for human fecal normalization in SARS-CoV-2 WBE surveillance was so far very limited and the most normalization process simply divides the concentration of viral load by the concentrations of PMMoV, and generate a unitless normalized values (Ai et al., 2021; D’Aoust et al., 2021; Feng et al., 2021; Greenwald et al., 2021; Isaksson et al., 2022). This study investigated several alternative chemical population biomarkers in SARS-CoV-2 WBE and normalization of SARS-CoV-2 to copies per capital. The concentrations of biomarkers determined by LC/MS/MS were applied to the exercise in correlation with population to generate their normalization coefficient. The viral loads per capita were determined using the normalization coefficient of each chemical population biomarker. Both direct and indirect approaches aimed at precisely estimating the population contributing to wastewater that would be applied in the following determination of the viral load per capita (Figs. 2 and 3). The normalization coefficient calculated from different biomarkers can be compared and evaluated before SARS-CoV-2 concentration was involved. Most importantly, our normalization approaches can be proceeded without daily flow volume and the size of the population using the developed regression functions (Tables S4 and S5), unlike the traditional normalization that requires the information of the daily flow volume and population size. The SARS-CoV-2 concentration was converted to mass using daily flow volume, followed by being divided by population served by the WWTP (Fig. 1A) to obtain viral loads per capita (copies/100,000 people).

In our normalization approaches, the parameter fold changes, the normalization coefficients (C1 and C2) standardized by C0 (from metadata), were utilized to evaluate the fitness of the biomarkers for each normalization approach as compared to the traditional method. The fold change that is closes to 1 indicates the highest accuracy. For example, in the direct approach, fold changes for CAF and PARA were 1.041±0.3111 (mean±standard deviation) and 1.057±0.389, respectively. In the indirectly approach, fold changes for CAF and PARA were 0.967±0.324 and 1.042±0.341, respectively. Both CAF and PARA showed high accuracy and low variability in either approach. On the contrary, 5-HIAA and PMMoV both showed significant difference in the values of the fold changes between two approaches. The 5-HIAA fold change was 1.150±0.661 with the direct approach but 1.470±1.144 in the indirect approach, whereas PMMoV performed better (1.003±0.586) with the indirect approach than (1.166±0.737) in the direct approach (Tables S6 and 7). The high accuracy and low variability for CAF and PARA are possibly attributed to high reproducibility of the analysis, high recovery rates, stability of these molecules, and low adsorption affinity to the solids fraction of wastewaters.

Furthermore, the regression functions established by CAF and PARA in our two approaches can be utilized to normalizing SARS-CoV-2 in the long-term monitoring without knowing daily flow volume and population size in the future WBE applications. The normalization approaches were validated using additional 64 samples collected from May 2021 (Table S1) with the established regression functions of CAF and PARA (Tables S4 and S5). The fold changes of CAF and PARA calculated from these additional 64 samples exhibited high precision and low variation in both direct and indirect approaches (Fig. 7). The results were consistent with our results from the developed models (Figs. 4 and 5), indicating the validation was successful.

The indirect normalization approach is more accurate than the direct approach. Both approaches were assumed the mass of biomarker correlates to the population, but the volume was taken into consideration in different ways. For the direct approach, the volume was utilized to calculate the ratio of population contributing to wastewater to the wastewater volume (Pv, number of people/L). It could vary due to different designs of WWTWs. The correlation between biomarker concentrations (μg/L) and Pp was significantly positive (Fig. 3). In the indirect approach, the flow volume was incorporated when the biomarker concentration was converted to mass. The correlation from both approaches remained positive even though flow volume was taken into consideration differently (Figs. 3 and 4). However, the correlation between biomarker mass and population was stronger as compared to the one between biomarker concentration and Pp, which affected the accuracy of population estimation (Tables S4 and S5). Subsequently, only the regression functions developed for indirect approaches were applied to estimate the real-time population.

This is the first study to normalize the SARS-CoV-2 load with biomarker estimated population and to accomplish viral load per capita with a universal unit – copies/person. Several previous studies utilized biomarker to normalize SARS-CoV-2 concentrations but only achieved a unitless results (eg. N1/N2 copies/copies of genetic biomarker). Green et al. reported the ratio of SARS-CoV-2:crAssphage in the wastewater; N1 or N2 copies/copies of biomarker (PMMoV, BCoV, HF183, crAssphage, and Bacteroides rRNA) in the wastewater were reported by several studies; D’Aoust et al. and Wolfe et al. presented copies/copy of PMMoV in solids (Table S10). Furthermore, Normalizing does not consistently improve correlations with COVID-19 cases within sewershed (Wolfe et al., 2021). For instance, Feng et al. reported normalizing SARS-CoV-2 to HF183 and PMMoV reduced correlations in 5 and 12 of 18 WWTWs respectively (Feng et al., 2021). Results from Greenwald et al. showed normalizing SARS-CoV-2 (N1) to PMMoV or Bacteroides rRNA did not improve he correlation (Greenwald et al., 2021). Ali et al. reported correlation was slightly improved by normalizing SARS-CoV-2 (N2) to PMMoV, but it was reduced by normalizing to crAssphage (Ai et al., 2021). Nevertheless, according to our findings (Fig. 9), the biomarker-estimated population should be incorporated into surveillance programs, so the normalization can reflect the real viral load per capita to be compared over time and cross facilities and be further utilized for predicting the trend of COVID-19 prevalence.

4.3. Relationship among estimated real-time population, SARS-CoV-2 in wastewater and COVID-19 incidences

No matter which biomarker was used to estimate the population, the population data from Census or WWTP operators could introduce significant possibility for error. Despite estimating real-time population was to examine its correlation with biomarker in the wastewater collected on Census Day (O’Brien et al., 2014), the Census population still did not capture the daily commuters or tourists. Consequently, the population provided by WWTP operators is the best available data for real-time population assessment.

The fluctuations in the population posed a challenge to WBE long-term monitoring (Polo et al., 2020). When the population contributing to the sewershed is expected to constantly change over the surveillance period (due to tourism, weekday commuters, temporary workers, etc.), population normalization becomes extremely critical for interpreting SARS-CoV-2 concentrations and predicting the trend and the infected population over time. We successfully demonstrated the utility of PARA for gauging small-area populations in real-time and captured the population dynamics in a college town and a tourist town (Fig. 5) since PARA gave the highest adjusted R-square with lowest MSE and 5-fold cross validation MSE in the population predicting model (Table 4). Our findings directly corresponded the fluctuations in the population due to seasonal activities in these tourist town and university
community, such as the summer breaks, holidays (e.g., Labor Day weekend in September) and tourism.

We strongly believe that population dynamic should be taken into consideration when the clinical cases are normalized for long-term monitoring. The CAF and its metabolites, PARA, have been proposed as anthropogenic markers to assess the population size and trace the discharge of domestic wastewater in rivers and lakes (Buerge et al., 2003). Senta et al. reported the PARA loads in the wastewater reflected the population dynamics (Senta et al., 2015). We demonstrated the significantly improved correlation between PARA-normalized SAR-Scov2 load per capita and the prevalence using a college town as an example (Fig. 9). Among 3 normalization scenarios (Fig. 9), only the PARA-normalized SAR-CoV-2 load (copies/person) and PARA-normalized cases (case number/100,000 people) yielded a statistically significant correlation (rho = 0.5878, p < 0.05). Our results indicated that a fixed population often derived from population census is not ideal for long term monitoring. It can be challenging to capture the population dynamic during the COVID-19 pandemic with the conventional methodologies based on periodic public surveys (such as census taking), augmented with a wide array of demographic statistics. Most of the inaccurate population data often derived from aged or incomplete sources such as census surveys or utility customers billed (e.g., Anderson et al., 2004; Banta-Green et al., 2009; Clara et al., 2011; Kasprzyk-Hordern et al., 2009; Neset et al., 2010; Ort et al., 2009; Rowse et al., 2010; Tsuzuki, 2006). Particularly during current pandemic, population dynamics often deviate significantly from the population estimated by the conventional methodologies due to the introduction of restrictions in control of the spread of SARS-CoV-2.

Unreliable population biomarkers often result in the poor correlation between the normalized SAR-CoV-2 levels and prevalence. For example, Feng et al. reported normalizing SAR-CoV-2 concentration in the wastewater to fecal marker HF183 and PMMoV reduced correlations in 5 and 8 of 12 WWTPs, respectively, compared to the correlation before normalization (Feng et al., 2021). Greenwald et al. also reported normalizing SAR-CoV-2 load using crAssphage, PMMoV, and Bacteroides rRNA in the wastewater samples deteriorated the correlation with daily case number per capita in comparison with the correlation between non-normalized concentrations and daily case numbers (Greenwald et al., 2021). According to our results, correlations could be significantly compromised when the fixed populations were applied to normalize the viral loads as well as the clinical cases.

5. Conclusion

Our findings suggested that the CAF metabolite, PARA, is a reliable population biomarker in SAR-CoV-2 wastewater-based epidemiology studies, due to its (1) better population indicators with higher accuracy, lower variability and higher temporal consistency as a population indicator to reflect the change in population dynamics and dilution in wastewater, (2) very limited exogenous sources, (3) high extraction efficiency with low variability in the extraction rates, (4) high stability, (5) resistance to chemicals in the wastewater, and (6) low sample volume requirement with simple sample preparation process. This chemical biomarker offers an excellent alternative to the currently CDC-recommended PMMoV genetic biomarker to help us understand the size, distribution, and dynamics of local populations for forecasting the prevalence of SARS-CoV2 within each sewershed. Furthermore, the regression functions embedded in the direct and indirect approaches of normalizing viral loads by biomarker could be applied to new data without known daily flow volume and population. Finally, the clinical cases should also be normalized by population dynamics when the correlation between SARS-CoV-2 and prevalence were examined.

Funding and acknowledgements

The authors would like to thank the Missouri Department of Health and Senior Services (DHSS) administering the funding. We would like to express our gratitude to the Missouri Department of Natural Resources (DNR) for coordinating the sample collection. Research reported in this publication was supported by funding from the Centers for Disease Control and the National Institute on Drug Abuse of the National Institutes of Health under award number U01DA053893-01. We would also like to thank the Center for Agroforestry at University of Missouri, USDA/ARS Dale Bumpers Small Farm Research Center under agreement number S8-6020-6-001 from the USDA Agricultural Research Service for supporting part of this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Centers for Disease Control or USDA-ARS.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.118985.

References

Agrawal, S., Oorschel, L., Lackner, S., 2021. Long-term monitoring of SAR-CoV-2 RNA in wastewater of the Frankfurt metropolitan area in Southern Germany. Sci. Rep. 11, 5372. https://doi.org/10.1038/s41598-021-84914-2. Ahmed, W., Angel, N., Edson, J., Bibby, K., Bindl, A., O’Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tschirke, B., Verhagen, R., Smith, W.J.M., Zaugu, 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. Sci. Total Environ. 728, 138764 https://doi.org/10.1016/j.scitotenv.2020.138764. Ahmed, W., Bertsch, P.M., Bibl, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gaylor, 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 SAR-CoV-2 from untreated wastewater. Sci. Total Environ. 739, 139960 https://doi.org/10.1016/j.scitotenv.2020.139960. Ai, Y., Davis, A., Jones, D., Lemeshow, S., Tu, H., He, F., Ra, P., Pan, X., Bohrerova, Z., Lee, J., 2021. Wastewater SAR-CoV-2 monitoring as a community-level COVID-19 trend tracker and variants in Ohio, United States. Sci. Total Environ. 801, 149757 https://doi.org/10.1016/j.scitotenv.2021.149757. Anderson, P.D., D’Aco, V.J., Shanahan, P., Chapra, S.C., Bueby, M.E., Cunningham, V.L., Duplissis, B.M., Hayes, E.P., Mastrocco, F.J., Parke, N.J., 2004. Screening analysis of human pharmaceutical compounds in US surface waters. Environ. Sci. Technol. 38, 838-849. https://doi.org/10.1021/es034330b. Armanious, A., Aeppli, M., Jacak, R., Refardt, D., Sigstam, T., Kohn, T., Sander, M., 2016. Viruses at solid-water interfaces: a systematic assessment of interactions driving adsorption. Environ. Sci. Technol. 50, 732-743. https://doi.org/10.1021/acs.est.5b04644. Asami, T., Katayama, H., Torrey, J.R., Visvanathan, C., Furuhashi, H., 2016. Evaluation of virus removal efficiency of coagulation-adsorption-inactivation and rapid sand filtration processes in a drinking water treatment plant in Bangkok, Thailand. Water Res. 101, 84–94. https://doi.org/10.1016/j.watres.2016.05.012. Banta-Green, C.J., Field, J.A., Chiaia, A.C., Sudakin, D.L., Power, L., De Montigny, L., 2009. The spatial epidemiology of cocaine, methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA) use: a demonstration using a population measure of community drug load derived from municipal wastewater. Addiction 104, 1874–1880. https://doi.org/10.1111/j.1360-0443.2009.02678.x. Barr Dana, B., Wilder Lynn, C., Caudill Samuel, P., Gonzalez Amanda, J., Needham Lance, L., Pirkle James, L., 2005. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. Environ. Health Perspect. 113, 192–200. https://doi.org/10.1289/ehp.7337. Brewer, A.J., Ort, C., Banta-Green, C.J., Berret, J.D., Field, J.A., 2012. Normalized diurnal and between-day trends in illicit and legal drug loads that account for changes in population. Environ. Sci. Technol. 46, 8305–8314. https://doi.org/10.1021/es202447r.
emissions—a conceptual approach combining substance-flow analysis with sampling analysis. Sci. Total Environ. 408, 3261–3269. https://doi.org/10.1016/j.scitotenv.2010.04.004.
Nuttall, K.L., Pingree, S.S., 1998. The incidence of elevations in urine 5-hydroxyindole-acetic acid. Model-Based Evaluation of Reduction Strategies for Micropollutants from Wastewater Treatment, pp. 167–174, p. 32.
O’Brien, J.W., Banks, A.P.W., Novic, A.J., Mueller, J.F., Jiang, G., Ort, C., Eaglesham, G., Yuan, Z., Thai, P.K., 2017. Impact of in-sewer degradation of pharmaceutical and personal care products (PPCPs) population markers on a population model. Environ. Sci. Technol. 51, 3816–3823. https://doi.org/10.1021/acs.est.5b02755.
O’Brien, J.W., Thai, P.K., Eaglesham, G., Ort, C., Scheidegger, A., Carter, S., Lai, F.Y., Mueller, J.F., 2014. A model to estimate the population contributing to the wastewater using samples collected on census day. Environ. Sci. Technol. 48, 517–525. https://doi.org/10.1021/es403251p.
Ort, C., Hollender, J., Schaerer, M., Siegrist, H., 2009. Model-based evaluation of reduction strategies for micropollutants from wastewater treatment plants in complex river networks. Environ. Sci. Technol. 43, 3241–3242. https://doi.org/10.1021/es902296v.
Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D.L., Singer, A.C., Graham, D.W., Ort, C., Hollender, J., Schaerer, M., Siegrist, H., 2009. Model-based evaluation of reduction strategies for micropollutants from wastewater treatment plants in complex river networks. Environ. Sci. Technol. 43, 3241–3242. https://doi.org/10.1021/es902296v.
Randazzo, W., Truchado, P., Cuebas-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 Res. 181, 115942 https://doi.org/10.1016/j.watres.2020.115942.
Refaeilzadeh, P., Yang, T., Liu, H., 2009. Cross-validation. Encycl. Database Syst. 5, 532–538. https://doi.org/10.1002/1-8499-7993-3.565-2.
Rico, M., Andrés-Costa, M.J., Pico, Y., 2017. Estimating population size in wastewater-based epidemiology. Valencia metropolitan area as a case study. J. Hazard. Mater. 323, 156–165. https://doi.org/10.1016/j.jhazmat.2016.05.079.
Robinson, C.A., Hsieh, H.-Y., Hsu, S.-Y., Wang, Y., Salcedo, B.T., Belenchia, A., Klutts, J., Zemmer, S., Reynolds, M., Semkiw, E., 2022. Defining biological and biophysical properties of SARS-CoV-2 genetic material in wastewater. Sci. Total Environ. 807, 150786 https://doi.org/10.1016/j.scitotenv.2021.150786.
Rosario, K., Symonds, E.M., Sinigalliano, C., Stewart, J., Breitbart, M., 2009. Pepper mild mottle virus as an indicator of fecal pollution. Appl. Environ. Microbiol. 75, 7261–7267. https://doi.org/10.1128/AEM.00410-09.
Rosiles-González, G., Ávila-Torres, G., Moreno-Valenzuela, O.A., Acosta-González, G., Leal-Bautista, R.M., Grimaldo-Hernández, C.D., Brown, J.K., Chávez-Quiroz, C., Betancourt, W.Q., Gerba, C.P., Hernández-Zepeda, C., 2017. Occurrence of pepper mild mottle virus (PMMoV) in groundwater from a karst aquifer system in the Yucatan Peninsula. Mexico. Food Environ Virol 9, 487–497. https://doi.org/10.1007/s12560-017-9309-1.
Rossell, V.F., Tangney, P., Hunt, C., Voulvoulis, N., 2010. Estimating levels of micropollutants in municipal wastewater. Water Air Soil Pollut 206, 357–368. https://doi.org/10.1007/s11270-009-0112-y.
Senta, I., Gracia-Lor, E., Borsotti, A., Zuccato, E., Castiglioni, S., 2015. Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. Water Res. 74, 23–33. https://doi.org/10.1016/j.watres.2015.02.002.
Sherchian, S.P., Shabir, S., Ward, I.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. Sci. Total Environ. 743, 140621 https://doi.org/10.1016/j.scitotenv.2020.140621.
State of Missouri, 2021. The Sewershed Surveillance Project [WWW Document]. ArcGIS StoreMaps. URL https://storymaps.arcgis.com/stories/7f5492486144da856d6c07f81aacf (accessed 9.22.21).
Summers, R.M., Mohanty, S.K., Gopishetty, S., Subramanian, M., 2015. Genetic characterization of caffeine degradation by bacteria and its potential applications. Microb. Biotechnol. 8, 369–378. https://doi.org/10.1111/1751-7915.12262.
Thai, P.K., O’Brien, J., Jiang, G., Gernjak, W., Yuan, Z., Eaglesham, G., Mueller, J.F., 2014. Degradability of creatinine under sewer conditions affects its potential to be used as biomarker in sewer epidemiology. Water Res. 55, 272–279. https://doi.org/10.1016/j.watres.2014.02.035.
Thai, P.K., O’Brien, J.W., Banks, A.P.W., Jiang, G., Gao, J., Choi, P.M., Yuan, Z., Mueller, J.F., 2019. Evaluating the in-sewer stability of three potential population biomarkers for application in wastewater-based epidemiology. Sci. Total Environ. 671, 248–253. https://doi.org/10.1016/j.scitotenv.2019.03.231.
Thomas, P.M., Foster, G.D., 2004. Determination of nonsteroidal anti-inflammatory drugs, caffeine, and triclosan in wastewateryby gas chromatography-mass spectrometry. J. Environ. Sci. Health Part A 39, 1969–1978. https://doi.org/10.1081/ESE-120039368.
Tszusui, Y., 2006. An index directly indicates land-based pollutant load contributions of domestic wastewater to the water pollution and its application. Sci. Total Environ. 370, 425–440. https://doi.org/10.1016/j.dtet.2006.07.014.
Udovičić, M., Bazdarevic, K., Bilic-Zulle, I., Petrovecki, M., 2007. What we need to know when calculating the coefficient of correlation? Biochem. Med. 17, 10–15. https://doi.org/10.11613/BM.2007.002.
WHO, n.d. Naming the coronavirus disease (COVID-19) and the virus that causes it [WWW Document]. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019-technical-guidance/naming-the-coronavirus-disease-covid-2019-and-the-virus-that-causes-it (accessed 7.14.21).
Wong, J.C.C., Tan, J., Lim, Y.X., Arivalan, S., Hapuarachchi, H.C., Mailepessov, D., Griffiths, J., Jayarajah, P., Setoh, Y.X., Tien, W.P., 2021. Non-intrusive wastewater surveillance for monitoring of a residential building for COVID-19 cases. Sci. Total Environ. 786, 147419 https://doi.org/10.1016/j.scitotenv.2021.147419.
Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W.L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghadi, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A.D., Erickson, T.B., Choi, P.R., Thompson, J., Ahn, E.J., 2020. SARS-CoV-2 Titors in wastewater are higher than expected from clinically confirmed cases. mSystems 5, 07f81aacf (accessed 9.22.21).
Wyss, M., Kaddurah-Daouk, R., 2000. Creatine and creatinine metabolism. Physiol. Rev. 80, 1107–1213. https://doi.org/10.1152/physrev.2000.80.2.1107.
Zhang, T., Breitbart, M., Lee, W.H., Run, J.-Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T., Rohwer, F., Ruan, Y., 2005. RNA Viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol. 4, e3. https://doi.org/10.1371/journal.pbio.0040005.