It is made available under a CC-BY-ND 4.0 International license .

Population Normalization in SARS-CoV-2 Wastewater-Based Epidemiology: Implications from Statewide Wastewater Monitoring in Missouri

3

- 4 Chenhui Li¹, Mohamed Bayati¹, Shu-Yu Hsu^{1,2}, Hsin-Yeh Hsieh¹, Wilfing Lindsi¹, Anthony
- 5 Belenchia³, Sally A. Zemmer⁴, Jessica Klutts⁴, Mary Samuelson⁴, Melissa Reynolds³, Elizabeth
- 6 Semkiw³, Hwei-Yiing Johnson³, Trevor Foley⁵, Chris G. Wieberg⁴, Jeff Wenzel³, Terri D.
- 7 Lyddon⁶, Mary LePique⁶, Clayton Rushford⁶, Braxton Salcedo⁶, Kara Young⁶, Madalyn
- 8 Graham⁶, Reinier Suarez⁶, Anarose Ford⁶, Dagmara S. Antkiewicz⁷, Kayley H. Janssen⁷, Martin
- 9 M. Shafer⁷, Marc C. Johnson⁶, Chung-Ho Lin^{1,2*}
- 10 *Corresponding Author:
- 11 Email address: linchu@missouri.edu
- 12

13 AFFILIATIONS

- ¹⁴ School of Natural Resources, University of Missouri, Columbia, MO 65211, USA.
- ² Center for Agroforestry, University of Missouri, Columbia, MO 65211, USA.

³ Bureau of Environmental Epidemiology, Division of Community and Public Health, Missouri
 Department of Health and Senior Services, Jefferson City, MO 65109, USA

- ⁴ Water Protection Program, Missouri Department of Natural Resources, Jefferson City, MO
 65101, USA
- 20 ⁵ Missouri Department of Corrections, Jefferson City, MO 65109, USA
- ⁶ Department of Molecular Microbiology and Immunology, University of Missouri, School of
 Medicine and the Christopher S. Bond Life Sciences Center, Columbia, MO 65211, USA.
- ⁷Wisconsin State Laboratory of Hygiene, University of Wisconsin Madison, Madison, WI
 53718, USA
- 25
- 26
- 27

It is made available under a CC-BY-ND 4.0 International license .

- 28 **KEYWORDS:** wastewater-based epidemiology; normalization biomarkers; real-time
- 29 population; caffeine, paraxanthine, pepper mild mottle virus; wastewater SARS-CoV-2
- 30 concentration; COVID-19 incidence rate
- 31

32 ABSTRACT



33

34 The primary objective of this study was to identify a universal wastewater biomarker for 35 population normalization for SARS-CoV-2 wastewater-based epidemiology (WBE). A total of 36 2,624 wastewater samples (41 weeks) were collected weekly during May 2021- April 2022 from 37 64 wastewater facilities across Missouri, U.S. Three wastewater biomarkers, caffeine and its 38 metabolite, paraxanthine, and pepper mild mottle virus (PMMoV), were compared for the 39 population normalization effectiveness for wastewater SARS-CoV-2 surveillance. Paraxanthine 40 had the lowest temporal variation and strongest relationship between population compared to 41 caffeine and PMMoV. This result was confirmed by data from ten different Wisconsin's 42 WWTPs with gradients in population sizes, indicating paraxanthine is a promising biomarker of the real-time population across a large geographical region. The estimated real-time population 43 44 was directly compared against the population patterns with human movement mobility data. Of 45 the three biomarkers, population normalization by paraxanthine significantly strengthened the 46 relationship between wastewater SARS-CoV-2 viral load and COVID-19 incidence rate the most 47 (40 out of 61 sewersheds). Caffeine could be a promising population biomarker for regions 48 where no significant exogenous caffeine sources (e.g., discharges from food industries) exist. In 49 contrast, PMMoV showed the highest variability over time, and therefore reduced the strength of the relationship between sewage SARS-CoV-2 viral load and the COVID-19 incidence rate, as 50 51 compared to wastewater data without population normalization and the population normalized by 52 either recent Census population or the population estimated based on the number of residential 53 connections and average household size for that municipality from the Census. Overall, the 54 findings of this long-term surveillance study concluded that the paraxanthine has the best 55 performance as a biomarker for population normalization for SARS-CoV-2 wastewater-based 56 epidemiology.

57 1. INTRODUCTION

Wastewater-based epidemiology (WBE) provides human activity information within
 sewershed boundaries by relating concentrations of chemical and biological "waste" materials in

It is made available under a CC-BY-ND 4.0 International license .

60 wastewater influent to population-scale use, consumption of pharmaceuticals, illicit drugs, or rates of exposure to industrial chemicals.¹ Over the last few years, WBE applications have been 61 62 emerging in infectious diseases or pathogens and antibiotic resistance, especially since the 63 COVID-19 pandemic. The WBE is widely recognized as a valuable tool for monitoring 64 community trends in COVID-19 with the advantages of providing an efficient and representative 65 population-pooled sample, and complementing community data, especially where timely COVID-19 clinical testing is underused or unavailable.² Moreover, WBE reveals the 66 67 underestimation of COVID-19 clinical testing because SARS-CoV-2 is shed by people with and without symptoms.^{3,4} The WBE can also provide a useful early warning of the emergence or re-68 69 emergence of COVID-19 in a community, and afford timely insights for public health 70 interventions, with previous studies showing that SARS-CoV-2 could be detected in wastewater 71 up to two weeks before the cases were reported.⁵ Furthermore, wastewater surveillance can be

72 implemented in most communities since municipal wastewater collection systems serve nearly

73 80 percent of U.S. households.⁶

74 The utility of WBE for cost-effective surveillance of SARS-CoV-2 levels in communities 75 was recognized early in the COVID-19 pandemic. However, substantive uncertainty remains in 76 how best to account for the contributing population and fecal strength. Robust population 77 biomarkers are necessary to determine the SARS-CoV-2 load per capita, such that 1) the changes 78 in wastewater virus concentration due to the dilution (e.g., increased volume resulting from 79 major rainfall events or receiving additional discharge from industrial or natural resources) and 80 2) population dynamics, are accounted for. Wastewater viral loads change with the variations in 81 daily wastewater flow, the proportion of industrial discharge, and the gross proportions of 82 solids, which can be influenced by the design and condition of the wastewater collection system.⁷ 83 For example, SARS-CoV-2 concentration is influenced if the WWTP is receiving wastewater 84 from a combined sewer system that collects domestic wastewater and rainwater runoff in the 85 same pipe; the weather effects, such as precipitation and infiltration/inflow into the sewers, 86 impact the human fecal concentration. In addition, the population size contributing to the 87 sewershed is expected to change over the surveillance period (due to deaths, births, tourism, weekday commuters, pandemic lockdown, temporary workers, etc.).¹ Therefore, temporal 88 89 variation in wastewater volume/strength and population size must be better accounted for to 90 make results more comparable over time.⁸ However, population normalization approaches are 91 still under development, and few studies have compared approaches/biomarkers systematically 92 for optimization of wastewater SARS-CoV-2 data for predicting the clinical prevalence of 93 COVID-19.9

94 The population variation is usually monitored and normalized using human fecal or urine 95 biomarkers. Suitable human population normalization controls should meet certain criteria.^{8,10,11} 96 For example, the chemical biomarkers should be specific to human metabolism, excreted into 97 sewage, exogenous sources are minimal, minimal intra- and inter-individual variance in daily 98 excretion, and levels in raw sewage well above the method detection limit. Furthermore, the 99 biomarkers must be stable in the wastewater for a reasonable long time (e.g., during the transport from the toilet to the sampling point and during sampling, storage, and analysis).¹² In addition, 100 101 there should be low variance in the per capita daily excretion and not be affected by 102 environmental variables such as season, weather, or geographic location.⁸

It is made available under a CC-BY-ND 4.0 International license .

103 Some commonly measured wastewater properties/chemicals, such as chemical and biological

- 104 oxygen demand, and total nitrogen and phosphorus, have been explored for wastewater
- 105 population indicators.^{13,14} The disadvantage of these environmental parameters is that they are
- 106 highly influenced by wastewater composition (i.e., industrial, domestic, or mixed) since they are 107 not only shed by humans but also from exogenous sources such as food waste processed by
- 107 not only shed by humans but also from exogenous sources such as food waste processed by 108 garbage disposals and fertilizer runoff.⁸ Ammonium originates from the breakdown of urea ¹⁵
- and is introduced via toilets and routinely measured by WWTP as a water quality parameter,
- 110 which is supposed to be less affected by non-human sources than chemical or biological oxygen
- demand and total phosphorous.¹⁶ However, Sweetapple et al. found that population
- normalization by orthophosphate and ammonium did not result in improvement of correlations
- between wastewater SARS-CoV-2 data and indicators of COVID-19 prevalence.¹⁴
- 114 In addition, a variety of endogenous and exogenous human biomarkers that can be measured 115 directly in wastewater samples have been evaluated to estimate their human fecal content. These
- markers include but are not limited to: (1) bacteria/viruses or molecules that are ubiquitous in
- human intestinal tracts, such as cross-assembly phage, human ribonuclease P,¹⁷ and *Bacteroides*
- 118 HF183:¹⁸ (2) an exogenous substance (or its' metabolite) after intentional consumption of a
- 119 substance (i.e., personal care products, food additives and dietary supplements such as
- 120 carbamazepine and gabapentin,¹⁹ artificial sweeteners,²⁰ and caffeine and its metabolite
- 121 (paraxanthine),¹⁰ pepper mild mottle virus (PMMoV);¹⁸ and (3) endogenous compounds that are
- 122 produced naturally in the body such as creatinine, cholesterol and its metabolite coprostanol,
- 123 cortisol, and serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA).¹¹
- 124 Pepper mild mottle virus (PMMoV), a virus ingested with pepper-containing food, has been 125 widely measured and frequently used for normalizing SARS-CoV-2 concentration data. The 126 PMMoV is one of the widely used normalization biomarkers in wastewater SARS-CoV-2 127 surveillance because it is believed to be present in high concentrations in wastewater, introduced 128 into the human body through the diet, and has the potential to serve as an RNA recovery control since it is a single-stranded RNA virus.^{21,22} D'Aoust et al. found PMMoV to be superior to 129 HF183 Bacteroides 16S ribosomal rRNA and eukaryotic 18S rRNA, as PMMoV showed more 130 reproducibility within and between WWTPs¹⁸. Creatinine, the endogenous nitrogenous waste 131 132 product, has been used to normalize the concentrations of other urinary excretion products to 133 account for urine dilution in clinical chemistry and is recommended as a possible biomarker for estimating the population.²³ The serotonin metabolite, 5-HIAA, has also been evaluated as a 134 135 wastewater marker, and was reported to be more stable within sewer systems than cortisol and 136 androstenedione.²⁴
- Caffeine (1,3,7-trimethylxanthine) is one of the world's most widely consumed dietary 137 ingredients, found in many globally popular products, including tea, cola and energy drinks, and 138 139 in some medications and nutritional supplements. Still, the most important source of this alkaloid 140 is coffee.²⁵ The excretion of the metabolite of caffeine, paraxanthine (1,7-dimethylxanthine), was 141 less affected by the genetic-based variation in pharmacokinetics than the parent compound was, therefore, suggested as a potential biomarker for dietary caffeine intake.²⁶ Chemicals (e.g., 142 143 paraxanthine) involved in endogenous metabolism (products of biosynthesis or catabolism) avoid 144 xenobiotics' problems for use as proxy measures for population since their association with per capita activities has higher fidelity.²³ Caffeine and paraxanthine are easy to detect due to the high 145 concentration levels (μ g/L) in untreated wastewater,²⁷ and are stable in wastewater samples 146

It is made available under a CC-BY-ND 4.0 International license .

- stored at 4° C,²⁸ which makes them ideal biomarkers. However, more research was needed for
- 148 caffeine and paraxanthine application in the concept of WBE.

In a previous study, we compared five biomarkers, PMMoV, creatinine, 5-HIAA, caffeine, and paraxanthine, based on two weeks' data, for their utility in normalizing SARS-CoV-2 loads and found caffeine and especially paraxanthine were the most reliable population biomarkers.²⁹ This study extended the investigation to the long-term weekly monitoring of 64 wastewater treatment plants (WWTPs) across Missouri for more than six months to compare the utility of caffeine, paraxanthine, and PMMoV as population biomarkers across seasons and distinct geographical areas with contrasting sewershed sizes.

156 **2. MATERIALS AND METHODS**

157 **2.1 Wastewater sampling and clinical COVID-19 case**

158 Triplicates of 50 mL of 24-hr composite influent (before primary treatment) samples were 159 collected once per week from 64 WWTPs in Missouri (Fig. 1), and a total of 41 weeks from May 160 2021 to April 2022 (weeks of 05/10/2021, 05/24/2021, 06/28/2021, and consecutively from 161 weeks of 07/19/2021 to 04/10/2202). A majority of samples were collected as 24-hr timeproportional composites and only 6 WWTPs were collected as 24-hr flow-proportional 162 composites. Each WWTP collected samples on the same day of the week during the study 163 164 period. The first half of the sampling period was dominated by the Delta variant, and the second half was dominated by the Omicron variant. In total, 2624 wastewater samples were collected. 165

166 Wastewater samples were transported in insulated shippers with ice packs to the University 167 of Missouri within 24 hr from the collection and then stored at 4°C until extraction. WWTPs reported their 24-hr flow rates and weekly new COVID-19 cases were provided by the Missouri 168 169 Department of Health and Senior Services (DHSS). Weekly clinical COVID-19 cases for each 170 WWTP service area were obtained by matching clinically confirmed COVID-19 case data (georeferenced using home address) to WWTPs sewershed boundaries in ArcGIS. Sewershed 171 172 boundaries were either provided by the municipality or, in many cases (12 WWTPs), the 173 municipality did not have geospatial data delineating their sewershed boundaries, so municipal 174 boundaries were used to approximate a service area.

175 The 64 WWTPs cover urban, semirural, and rural locations throughout Missouri with the sewershed population ranging from 900 to 490,000. Sewershed size (metadata population) was 176 177 either provided by the WWTP or estimated based on the number of residential connections 178 reported in a WWTP's discharge permit and the average household size for that municipality 179 from the most recent U.S. Census at the time the facility began sampling. Based on the 180 population served by the 64 WWTPs, there is the potential to monitor 50% of the 6.15 million Missouri population.³⁰ A summary of each facility, including the population served and the 181 182 locations, is provided in Table S1. Additionally, ten wastewater composite samples collected 183 from WWTPs in Wisconsin during the week of 06/07/2021 were utilized as a data set for model 184 evaluation and validation.

It is made available under a CC-BY-ND 4.0 International license .

185 2.2 Mobility data

Daily data on mobility (driving and walking) were downloaded for Columbia, MO,
between 01/13/2020 and 04/13/2022 from Apple Mobility Trends Reports³¹. Apple mobility data
has no demographic information about the users. The original mobility indices were relative
percentages to the reference date on 01/13/20, which were scaled to the maximum observed
during the study period.

191 **2.3.** Wastewater concentrations of caffeine and paraxanthine.

192 2.3.1. Extraction of caffeine and paraxanthine

For all 2,624 wastewater samples, 1.5 ml was centrifuged at 10000 rpm for 10 minutes, and 193 194 0.75 ml supernatant was extracted and mixed with 0.75 ml ammonium acetate buffer (10 mM 195 ammonium acetate and 0.1% formic acid in water). Then, 10 µL of formic acid was added to precipitate the large molecules in the wastewater, followed by spiking Caffeine-C¹³ to evaluate 196 197 the recovery of endogenous caffeine. For improved sample storage stability, the 0.75 ml 198 ammonium acetate buffer was replaced by 0.75 ml 100% methanol starting from December 199 2021. The two preparation methods (ammonium acetate buffer and methanol) were compared, 200 and no significant difference in caffeine and paraxanthine concentrations was found before the 201 transition (Table S2). Finally, the mixture was filtered through 0.2 µm PTFE filters (13mm) 202 (Waters, USA) before the liquid chromatography with tandem mass spectrometry (LC-MS/MS)

analysis.

204 2.3.2. Liquid chromatography-tendon mass spectrometry analysis.

205 The quantification of caffeine and paraxanthine was performed on a Waters Alliance 2695 206 High Performance Liquid Chromatography (HPLC) system coupled with Waters Acquity TQ 207 triple quadrupole mass spectrometer (MS/MS). The analytes were separated on a Phenomenex 208 (Torrance, CA) Kinetex C18 (100mm x 4.6 mm; 2.6 µm particle size) reverse-phase column. 209 The mobile phase consisted of 10 mM ammonium acetate and 0.1% formic acid in water (A) and 210 100% acetonitrile (B). The gradient conditions were 0 - 0.3 min, 2% B; 0.3-7.27 min, 2-80% B; 211 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 a 212 flow rate of 0.5 mL/min. The ion source in the MS/MS system was electrospray ionization (E.I.) 213 operated in positive ion mode with a capillary voltage of 1.5 kV. The temperature of the 214 ionization source was 150°C, and that of the desolvation zone, 450°C. The optimized collision 215 energy, cone voltage, and molecular and product ions of the biomarkers are summarized in Table 216 1.

217 2.4.1 Wastewater concentrations of SARS-CoV-2 and PMMoV

218 2.4.1. RNA extraction

Samples from the weeks of 09/13/2021 to 04/10/2022 were processed for both SARS-CoV-2 and PMMoV quantification. First, 50-ml raw wastewater samples were spun at 2000 ×g for 5 min to remove large particulates, then vacuum filtered through a 0.22 µm filter (Millipore cat#

222 SCGPOO525). Then, 35.5 mL of filtered wastewater was mixed with 12 mL of 50% (W/V)

It is made available under a CC-BY-ND 4.0 International license .

223 polyethylene glycol (PEG, Research Products International, cat# P48080) and 1.2M NaCl,

followed by equilibration for 2 h at 4°C, all done on the day of sample receipt. Afterwards,

samples were centrifuged at 12,000 x g for 2 h. RNA was extracted from the pellet using the

- 226 Qiagen Viral RNA extraction kit following the manufacturer's instructions after removing the
- supernatant. RNA was eluted in a final volume of 56 μ L and stored at -80°C if it was not
- 228 processed immediately.

229 2.4.2. RT-qPCR

The extracted RNA was used to perform RT-qPCR quantification of the genetic material of
 SARS-CoV-2 and PMMoV, separately. SARS-CoV-2 was quantified using the primer and
 control sets described in Robinson et al.,³² and PMMoV was quantified using the primer sets
 described in Hsu et al.²⁹ A plasmid carrying a PMMoV gene along with an N gene fragment was
 constructed, purified from *Escherichia coli*, and used as standards for the RT-qPCR assay. Final
 RT-qPCR one-step mixtures consisted of 5 µL TaqPath 1-step RT-qPCR Master Mix (Thermo

Fisher), 500 nM of each primer, 125 nM of each TaqMan probes, 5 µl of wastewater RNA

- extract, and RNase/DNase-free water to reach a final volume of $20 \,\mu$ L. All RT-qPCR assays
- 238 were run in duplicate on a 7500 Fast real-time qPCR instrument (Applied Biosystems). The

reactions were initiated with one cycle of UNG incubation at 25°C for 2 min and then one cycle

of reverse transcription at 50°C for 15 min, followed by one cycle of activation of DNA

241 polymerase at 95°C for 2 min and then 45 cycles of 95°C for 3 sec for DNA denaturation and

55°C for 30 sec for annealing and extension. The data would be collected at the step of 55°C

extensions. The concentration of SARS-CoV-2 and PMMoV genomes in each sample, $(N/L)^{21} = ar [D)(Ma)^{21} (applied(L)) successful to d using Eq. 1$

244 [*N1.N2*]_{SARS} or [PMMoV] (copies/L), was calculated using Eq. 1.

245 $[N1.N2]_{SARS}$ or $[PMMoV] = N/(V_{RT-qPCR} \times (V_{sample} / (V_{extracted} / recovery rate))) (1)$

246 where N (copies/reaction) is the gene copies detected in each RT-qPCR reaction, V_{RT-qPCR} is the

247 volume of RNA used for RT-qPCR (5 μ L), V_{sample} is the wastewater sample volume initially

248 used for the concentration step (35.5 mL), and V_{extracted} is the total volume of nucleic acid

extracted (56 μ L).

250 **2.5 Statistical analyses**

251 The statistical analyses included the following three steps: 1) estimation of the real-time

252 population using the wastewater biomarker mass loads; 2) population normalization for

253 wastewater SARS-CoV-2 viral load and COVID-19 case to wastewater SARS-CoV-2 viral

concentration (copies/week/10,000 people) and COVID-19 incidence rate (case/week/10,000

people), and 3) comparison of the strength of the relationships between population normalized
 wastewater SARS-CoV-2 viral load and COVID-19 incidence rate under different normalization

257 scenarios.

258 2.5.1. Real-time population estimation using wastewater biomarker loads

All the data analyses were conducted using the R program.³³ To compare biomarkers' variability and temporal consistency, the coefficient of variation (CV%) of the three biomarkers across seven months of data at each of the 64 WWTPs was calculated. Only 7-months of data

It is made available under a CC-BY-ND 4.0 International license .

262 (n=1596) between the weeks of 09/13/2021 to 04/10/2022 were included in the analyses to
 263 compare three biomarkers since PMMoV was only measured after 09/13/2021. The relationship

between weekly biomarker load and the population contributing to the wastewater was examined

using linear regression. The biomarker load of biomarker *i* for *j* WWTP, B_{ij} , was calculated as

266 $B_{ij} = [B]_{ij} \times F_j \times 7 \times 3.7841 \times 10^6$ (2)

in which $[B]_{ij}$, the biomarker *i* concentration in *j* WWTP wastewater sewershed, was determined by LC-MSMS or RT-qPCR. F_i is the wastewater daily flow volume (million gallons per day) for

WWTP_{*j*}. Constants 3.78541 and 7 are applied to convert the imperial unit to metric unit and daily to weekly biomarker load, respectively.

270 duriy to weekly biomarker foud, respectively.

- 271 The linear regression model was conducted as
- $B_i = \beta_i P + \epsilon_i \tag{3}$

Where *P* is the corresponding population from metadata data. The modeling was based on the assumption that the mass load entering the treatment plant per day of a chemical is proportional

to the contributing population. Since the actual real-time population was not available, the best

estimation of it was the metadata population. B_i is the biomarker mass load, ϵ_i the error term, and

277 β_i the estimated parameter for biomarker *i*. Log transformation was applied to the population and

biomarker load.

279 In addition, ten wastewater samples (each from different WWTFs) from Wisconsin collected

during the week of 06/07/2021 were used to validate equations (3) using Root Mean Square

281 Error (RMSE). The RMSE is the standard deviation of the prediction errors.

Then, the real-time population P_{RT} by biomarker *i* was estimated using all three biomarkers according to the equation

284

 $P_{RTi} = (B_i - \epsilon_i) / \beta_i \tag{4}$

285 2.5.2. Population normalization of wastewater SARS-CoV-2 load and COVID-19 incidence 286 rate.

Wastewater SARS-CoV-2 load was then normalized to population (copies/week/10,000
people) by dividing the SARS-CoV-2 load per week by each of the three estimated population
metrics:

290 Normalized
$$[SARS-CoV-2]_i = \frac{[N1.N2]_{SARS} \times (F_j \times 3.7841 \times 10^6) \times 7 \times 10000}{Real-time Population_i or P_{metadata}}$$
 (5)

291 in which $[N1.N2]_{SARS}$ (copies/L), the SARS-CoV-2 concentration in the wastewater sewershed, 292 was determined by RT-qPCR using equation (1). The COVID-19 incidence rate was defined as: 293 294 Normalized Covid-19 incidence rate_i = $\frac{Clinic case number \times 10000}{Real-time Population_i or P_{metadata}}$ (6) 205 Perulation permulsion better allows the results to be compared percess WWTPs that correct

Population normalization better allows the results to be compared across WWTPs that servedifferent size populations.

It is made available under a CC-BY-ND 4.0 International license .

297 2.5.3. Effectiveness of population normalization on the relationships between wastewater 298 SARS-CoV-2 load and clinical COVID-19 incidence rate.

299 To compare the outcomes of population normalization of different biomarkers, the strength 300 of the linear regression models from normalized wastewater SARS-CoV-2 load to normalized 301 COVID-19 incidence rate were compared across markers, as well as when there was no 302 population normalization (i.e., the correlation strength of raw viral load per week with total case 303 number per week). Specifically, the goodness of fit of linear regression models was compared 304 within each WWTP using the coefficient of determination (\mathbb{R}^2) , the measure of "variance 305 explained". Analysis of variance (ANOVA) and pairwise comparisons were conducted on the R^2 306 of four groups (without normalization, metadata population normalization, caffeine estimated 307 population normalization, paraxanthine estimated population normalization, and PMMoV 308 estimated population normalization) within each WWTP. Log transformation was applied to both 309 wastewater SARS-CoV-2 load and the COVID-19 incidence rate. Since there are many "0" 310 COVID-19 cases, $log_{10}(x+1)$ was employed for the COVID-19 incidence rate.

311 3. RESULTS

312 **3.1. Temporal variations of biomarkers**

313 The coefficient of variations (CV%) of the three biomarkers across seasons at each of the 314 64 WWTPs was calculated (Fig. 2). The mean CV% across all the facilities were not different 315 between caffeine (43%) and paraxanthine (40%), but both significantly lower than the CV% of 316 PMMoV (mean=67%) (p<0.001). Paraxanthine was the only biomarker with all the CV% lower 317 than 100% and the smallest range of CV% among 64 WWTPs (caffeine = 263%, paraxanthine = 318 54%, and PMMoV = 154%). The lower CV% means of caffeine and paraxanthine than PMMoV 319 indicated that caffeine and paraxanthine changed less over time than PMMoV. No significant 320 relationship between each biomarker's CV% and population size was found (data not shown). Compared to caffeine and PMMoV, the smallest CV% range of paraxanthine among WWTPs 321 322 showed paraxanthine is less influenced by population attributes (i.e., population size, the 323 composition of age, race, and ethnicity, and environmental factors such as season and fecal 324 strength across geographic regions).

325 The outlier of CV% of caffeine (280%) was the Sikeston Wastewater Treatment Plant 326 (SKSTN), which is nearby a global industry factory, Unilever ice cream, manufacturing the 327 distributed worldwide Magnum chocolate ice cream bar (caffeine concentration in chocolate is 328 around 420 μ g/g). Eleven out of 16 samples with caffeine concentrations over 200 μ g/L/week 329 were from SKSTN. The extreme caffeine concentrations were not likely all shed by humans 330 from such a small town with a population of 17,000 but rather from industrial waste. After 331 removing samples from SKSTN, there was a strong linear relationship between caffeine and 332 paraxanthine concentration (Fig. S1). Therefore, caffeine is still a reliable biomarker for the area 333 where no such exogenous caffeine sources exist.

334 **3.2. Real-time population prediction by biomarkers**

Linear regression models were established for biomarker mass load per week and population sizes obtained from the metadata for both Missouri. Then, real-time population (P_{RT})

It is made available under a CC-BY-ND 4.0 International license .

337 was predicted for each WWTP with all available data. The estimated P_{RT} of tourism town

338 (ANON2) from all three biomarkers was generally higher than the metadata population, which is

reasonable since the metadata population primarily only counts residents. The estimated P_{RT} for

the large metropolitan area, Kansas City Blue River, was almost always lower than the metadata

population (Fig. S3A), probably indicating a population decline in the past two years since the

census survey in 2020. There was a strong relationship between the metadata population and the
 total case number during the seven months from 09/13/2021 to 04/10/2022 (Fig. S3B and S3C).

Kansas City Blue River was the "outlier" in Fig S3B, suggesting that the PRT of the Kansas City

345 Blue River area might be substantially lower than the metadata population. Upon further

346 investigation, it was learned that KCBLU WWTP was in the process of upgrading and the served

347 area changed periodically compared to the metadata population that the facility initially provided

348 (personal communication).

349 Therefore, models between biomarker mass load and population were revised by 350 removing KCBLU WWTP in the models for Missouri (Fig. 3A, 3C, 3E) and metadata 351 normalization data in KCBLU were also removed in the following analyses. Linear regression 352 models were also established for biomarker mass load per week and population sizes obtained 353 from the metadata for Wisconsin (Fig. 3B, 3D, 3F). Within Missouri, the model with paraxanthine had the highest R² among all three biomarkers. Similar trends were observed in 354 355 Wisconsin models, especially the paraxanthine model, which confirmed that the relationship 356 between biomarkers and population could be applied beyond Missouri. In addition, Wisconsin 357 samples were used to test Missouri models (Fig. 3A, 3C, 3E) using RMSE, which were 0.24, 358 0.19, and 0.87 for models of caffeine, paraxanthine, and PMMoV. The lowest RMSE of the 359 paraxanthine model indicated it is the best predictor of the population too.

The ratios of the estimated real-time population to the metadata population (P_{RT} / $P_{METADATA}$) by paraxanthine had the smallest mean and standard deviation among the 64 WWTPs (Table 2). Among the three biomarkers, caffeine and paraxanthine, in general, are closer to the metadata populations. In contrast, estimated P_{RT} from PMMoV wavered with time dramatically. The smaller variations of estimated P_{RT} from caffeine and paraxanthine than PMMoV are consistent with the low temporal variations of caffeine and paraxanthine relative to PMMoV.

367 A case study of validation of biomarker population estimation using the Apple mobility 368 data was conducted in a college town, Columbia, Missouri. The Apple mobility showed that 369 demographic migration was influenced by the COVID-19 pandemic and the school events in 370 Columbia, Missouri (Fig. 4A). For instance, there was a sharp decline of mobility in April 2020 371 after the Missouri "Stay home" order. Both mobility data and the estimated P_{RT} estimated by paraxanthine showed the fluctuations of the population with school semesters and holidays (Fig. 372 373 4A and B). For instance, Labor Day weekend, Thanksgiving break, and winter break had lower 374 P_{RT} and mobility than other times since a large group of students traveled back to their 375 hometowns during these holidays. Also, the week of the University of Missouri Homecoming 376 event (10/05/2021) had a P_{RT} and mobility increase.

It is made available under a CC-BY-ND 4.0 International license .

377 **3.3. Effectiveness of population normalization of different wastewater biomarkers**

378 The strengths of the relationships between the population normalized SARS-CoV-2 RNA 379 load and COVID-19 incidence rate were compared within each WWTP (Table 3). A total of 59 380 out of 64 WWTPs showed a significant positive relationship between wastewater SARS-CoV-2 381 RNA load and clinical COVID-19 incidence rate for all population normalization scenarios. In 382 addition, models were significant except for the PMMoV estimated population normalization 383 model for two WWTPs (ANON1 and MACON). Within 57 of 64 WWTPs, biomarker 384 normalizations strengthened the relationship between SARS-CoV-2 RNA load and normalized COVID-19 incidence rate (i.e., R^2 increased). The mean R^2 decreased (α level=0.1) in the order: 385 386 paraxanthine estimated real-time population normalization > without population normalization 387 and metadata population normalization > caffeine estimated real-time population

- 388 normalization > PMMoV estimated real-time population normalization, indicating that
- 389 paraxanthine is the best population normalization biomarker (Fig. S3).

390 The regression models of viral loads with clinical incidence after normalization to

391 paraxanthine estimated population had the highest R^2 among all five models for 40 out of 61

392 WWTPs (Fig. 5). In contrast, the strength of the relationship between wastewater viral load and

393 clinical incidence after PMMoV and caffeine estimated population normalization became weaker

than without population normalization or with metadata population normalization (Fig. S3).

395 Therefore, the time series of wastewater viral load and COVID-19 incidence rate with

normalization of paraxanthine estimated population was plotted for each WWTP (Fig. 6).

Weekly wastewater viral copies followed similar patterns with clinical incidence rates for most
 WWTPs. Furthermore, the strength of the relationship between wastewater SARS-CoV-2 viral

398 WWTPs. Furthermore, the strength of the relationship between wastewater SARS-CoV-2 viral 399 load and the COVID-19 incidence rate was significantly increased with increasing cases and

400 metadata population (Fig. 7A and 7B).

401 **3.4. Limit of detection of wastewater SARS-CoV-2 surveillance**

The linear regression relationship between wastewater SARS-CoV-2 RNA load with clinical
 cases was tested for models without population normalization and paraxanthine population
 normalization (Fig. S4). According to the paraxanthine population normalization model, the
 process limit of detection (PLOD) (entire process from the sampling to the analysis of RT-qPCR)
 in the wastewater over should be:

407
$$PLOD1 (copy/week) = 1.3 \times 10^{11}$$
 (7)

A validation calculation was conducted using the PLOD (<3,954 GC/50ml; 95% probability of
 detection) of the US CDC N1 RT-dPCR and RT-qPCR assays from a study based on wastewater
 SARS-CoV-2 seeding experiment by Ahmed et al.³⁴ (Table 4):

411
$$PLOD2 \text{ (copy/week)} = (3,954 \text{ copy}/50 \text{ ml}) \times 1000 \text{ ml} \times F_i \times 3.7841 \times 10^6$$
 (8)

412 Although Ahmed et al. used a different adsorption extraction method of virus to evaluate N1

413 copy instead of average copy of N1 and N2 as in this study, both PLOD1 and PLOD2 showed

414 that the N1N2 mass load was higher than PLODs for ANON1 (the smallest WWTP) during the

415 three weeks when the main outbreak occurred. The PLOD1 was likely higher than the actual

It is made available under a CC-BY-ND 4.0 International license .

416 value since the clinical cases are likely less than the real cases due to the asymptomatic and 417 underreported cases.

418 **4. DISCUSSION**

419 **4.1 Paraxanthine was the optimal population biomarker**

420 The reduced strength of the relationship between the PMMoV-population normalized SARS-CoV-2 load and COVID-19 incidence rate was consistent with previous studies that showed 421 422 PMMoV had mixed or adverse effects on the correlation between wastewater measurements of SARS-CoV-2 and clinical cases.^{35–37} To the best of our knowledge, there is only one nationwide 423 study to date (conducted by Biobot Analytics) that is comparable with our study in terms of the 424 425 temporal and spatial magnitudes of sampling. That study collected 2,433 samples from 55 426 WWTPs across the U.S. and found that PMMoV normalization did not always improve the 427 correlation between wastewater measurement and clinical cases.³⁵ Feng et al. showed that PMMoV normalizations reduced the correlations between SARS-CoV-2 concentration and 428 429 COVID-19 incidence for 8 of 12 WWTPs and suggested that variability's influence across 430 measurement for human viral is stronger than that of differences in the fecal loads in the 431 samples³⁶. For some sewersheds where the normalizations by metadata population and without 432 population normalization were better than the normalizations by caffeine and PMMoV estimated 433 population, probably because most WWTPs serve rural areas in Missouri where the population 434 does not fluctuate with time. However, each WWTP was weighted equally in this analysis 435 regardless of the population size it serves. Furthermore, many people worked from home during 436 the pandemic; therefore, the real-time population was expected to be closer than usual to the 437 metadata population for many regions.

438 The intensified relationship between wastewater SARS-CoV-2 viral load with clinical 439 COVID-19 cases by wastewater paraxanthine concentration for 2/3 of the WWTPs and the 440 consistency of the relationship between wastewater paraxanthine concentration and population 441 between Missouri and Wisconsin demonstrated that paraxanthine is a reliable population 442 biomarker across large geographical regions with different sizes of the population. The superior 443 performance of paraxanthine over PMMoV could be attributed to its 1) much longer half-life, 2) 444 less exogenous sources and variability of excretion rate intra-individual and inter-individual, 3) 445 easier and more accurately determined in the wastewater.

446 4.1.1 Stability

447 A promising population normalization biomarker should be stable in wastewater. 448 Caffeine and its metabolites in untreated wastewater were stable during 24 hr storage at 4°C and 20°C and also stable at -20°C for up to four weeks.²⁸ One study showed PMMoV was stable for 449 21 days at 4°C, 25°C, and 37°C, ³⁸ However, it was mostly based on different water types that had 450 451 simpler contents than wastewater, such as deionized water and autoclaved wetland water. In 452 addition, the conclusion that PMMoV has high stability over time was drawn from much shorter 453 periods of investigations or a small number of samples collected from fewer WWTPs than in our 454 study^{4,18,21,22,36}. For instance, a highly cited article from Kitajima et al. found that PMMoV did 455 not change seasonally, but was only based on 48 samples collected from two WWTPs over 12 months.²¹ Wu et al. only tested PMMoV stability for SARS-CoV-2 fecal biomarker on samples 456

It is made available under a CC-BY-ND 4.0 International license .

457 collected over one week during March 2020.⁴ D'Aoust et al. found that PMMoV is a better

458 normalization biomarker than *Bacteroides* 16S rRNA or human 18S rRNA, which was also only

459 based on 23 samples from two WWTPs.¹⁸ In contrast, higher stability of paraxanthine than

460 PMMoV in our study was based on 1569 samples from September 2021 to April 2022, which

461 covered three months of the Delta variant in 2021 and the major spike of the Omicron variant at

the beginning of 2022. The 64 WWTPs represent 50% of the entire Missouri population, from

463 metropolitan areas such as St. Louis to small rural towns with only 900 people.

464 *4.1.2. Sources*

A reliable population biomarker should primarily be shed by humans and respond to the 465 population size, not environmental factors.⁸ A regular and constant consumption is a further 466 prerequisite for a good marker.³⁹ Loads of caffeine in untreated wastewater reflect not only 467 consumption, metabolism, and excretion of the compound but also caffeine from beverages, 468 foods, and pharmacologic that were poured out directly.⁴⁰ Caffeine is transformed in the human 469 470 liver into more than 20 metabolites, primarily dimethylxanthines (paraxanthine, theobromine, and theophylline), dimethyl- and monomethyluric acids, and uracil derivatives³⁹. Between 0.5% 471 and 10% of the caffeine in human body is excreted unmetabolized via urine.³⁹ Exogenous 472 473 sources from industrial wastewater can influence caffeine if the WWTP is a combined sewer 474 system, which also has potential pollution issues from the outflow.⁷ Besides, we also had several 475 samples that could not detect caffeine and paraxanthine after heavy storms (data removed from 476 analyses). Therefore, if combined sewer systems are chosen for WWTP, individuals should be 477 aware that in some circumstances, like high rain events, readings of viral load and other 478 measurements may need be corrected for dilution.

479 As a caffeine metabolite, paraxanthine mainly comes from human metabolism.⁸ Therefore, 480 paraxanthine relates to the population better than caffeine. Among the metabolites of caffeine, paraxanthine is the most abundant caffeine metabolite in wastewater.²⁶ In addition. 1-481 methylxanthine, 7-methylxanthine, and 1,7dimethyluric acid are also metabolites of theophylline 482 483 and theobromine respectively, which are present in some foods, drinks, and pharmaceutical formulations.³⁹ Paraxanthine was believed to be the optimal biomarker of caffeine intake as a 484 485 population biomarker.¹ The more minor variations of paraxanthine among different WWTPs 486 were probably due to similar levels of caffeine intake among different groups in the population. 487 The main factor driving paraxanthine load in the wastewater was the coffee consumption rate, 488 which can be influenced by the composition of the population age since kids do not drink as 489 much as adults. In addition, average consumption is 70 mg per person per day but varies in the different countries due to the different culture.³⁹ However, caffeine is one of the most widely 490 491 consumed dietary ingredients worldwide, and thus paraxanthine as its metabolite has the 492 potential to be used as a population biomarker worldwide.

In comparison, PMMoV is a pepper pathogen virus that often is found in human feces, as well as peppers and processed pepper products from all over the world, such as dry spices and sauces.⁴² Unlike caffeine or tea are widely consumed by people all over the world, the importance of pepper in cuisine varies depending on the regions. For example, chili pepper is the defining ingredient of New Mexican food but not for most European cuisine. Therefore, the population race composition might influence the pepper's consumption and thus PMMoV concentration. The detection level of PMMoV in human feces varies greatly from 7% to 95%,

It is made available under a CC-BY-ND 4.0 International license .

500 depending on the study's regions and also between adults and children, even in the same

501 regions.⁴³ The PMMoV concentration in our study had an average of $1.5 * 10^8 \text{ copy/L}$, which is

502 consistent with the previous study that showed the PMMoV concentration from raw wastewater

ranged from 10^{8} - 10^{10} copy/L,⁴⁴ but PMMoV in our study largely ranged from $3*10^{4}$ to $2*10^{9}$ 504 copy/L.

505 In addition, SARS-CoV-2 RNA and PMMoV are shed in fecal, but caffeine and paraxanthine 506 are discharged through urine. Humans urinate much more frequently than bowl movement, likely 507 contributing to less variance in quantity among individuals than bowel movement, which 508 possibly is one of the reasons that biomarkers in urine such as caffeine and paraxanthine had less

509 variations and are more representative of population size.

510 4.1.3. Quantification

511 Being easy to be determined with high repeatability is another criterion of a good population 512 marker. The low variation of caffeine and paraxanthine owes to the high analytical sensitivity of 513 LC-MSMS system and the consistent and high extraction recovery rate. With the development of 514 technology, most of LC-MSMS systems have ng/mL to pg/mL level sensitivity. The instrumental 515 detection limits for caffeine and paraxanthine were reported as 1.4 and 2.4 pg/injected, and instrumental quantification limits for caffeine and paraxanthine were 3.6 and 6.6 ng/L using 516 517 API5500 QqQ equipped with a Turbo Ion Spray source.²⁸ The sensitivity of our HPCL-MSMS 518 with electrospray ionization is at ng/mL, but it is sufficient for wastewater caffeine and 519 paraxanthine quantification. The caffeine concentration in wastewater influent water was 520 reported from 20-300 µg/L (Canada), 20 µg/L (U.S.), and 147±76 µg/L (Germany).²⁸ In our 521 study, the average caffeine and paraxanthine concentrations were 71 ug/L and 17.5 ug/L, 522 respectively. The recoveries of caffeine and paraxanthine from untreated wastewater were 88% and 76% during the similar storage temperature and extraction method to our study.²⁸ Our 523 524 previous study showed that the recovery rates of caffeine and paraxanthine during injection are 525 $101\pm7\%$ and $92\pm3\%$ ²⁹ In addition, the repeatability for caffeine and paraxanthine was quite high (CV%: 12% and 5%).²⁸ 526

527 Unlike the simple chemical analysis on LC-MSMS, PMMoV measurement has a very

complex workflow. First, wastewater samples require the application of concentration steps
 before extraction of the RNA fragments. Then, highly-sensitive molecular assays using RT-

- gPCR or RT-dPCR (digital PCR) will be applied to quantify the PMMoV concentration.³⁴
- 531 Consequently, there are many factors that may contribute to the large variation, such as the
- 631 Consequently, there are many factors that may control te to the large variation, such as the
 632 efficiency of primary concentration, loss through nucleic acid extraction, and inhibition of
- reverse transcription or PCR amplification. PMMoV's recovery rates reported in previous studies
- varied but were generally lower than caffeine and paraxanthine (e.g., 45±26% using direct
- extraction and only >10% using electronegative (H.A.) filters).^{36,45} The recovery rate of PMMoV
- 536 in this study was not tested directly, but the virus concentration methods (PEG concentration) in
- 537 this study preserved SARS-CoV-2 N1N2 at approximately 62% signal on average and 2.7 times
- 538 higher for Puro.³² The PMMoV's process limits of quantification and detection were evaluated in
- 539 diluted wastewater in the coastal water,⁴⁶ which is different from PMMoV in raw wastewater.

It is made available under a CC-BY-ND 4.0 International license .

540 4.2 Normalization approach recommendation for SARS-CoV-2 wastewater-based 541 epidemiology

542 A previous study showed that outbreaks could be detected in buildings with as many as 1,000 543 occupants.⁴⁷ This study showed that the wastewater could also detect community-level outbreaks with a small population size, which revealed that the process limit of detection for wastewater 544 545 surveillance (e.g., the fewest infections in a community that can be reliably detected in 546 wastewater) is quite low ($\sim 0.1\%$ of the population, 1-2 cases in 900) (Table 4). However, the 547 PLOD could be overestimated since the actual cases could be higher than just 1 or 2 cases. The 548 smallest WWTP, ANON1, still showed a significant relationship between the wastewater viral 549 copies and the COVID-19 incidence rate, except when using the normalization of PMMoV 550 estimated population (Table 3). Our previous study found that approximately 20% of the tested 551 WWTPs in Missouri, U.S., receive some input from industries, possibly discharging some chemicals that suppress the viral signals in wastewater.⁴⁸ MACON WWTP was one of the 552 examples.⁴⁸ However, positive relationship between the wastewater viral load and the COVID-553 554 19 incidence rate for MACON except for PMMoV estimated population normalization (Table 3). 555 Therefore, WBE is a feasible approach for population-level monitoring of COVID-19 disease. 556 PMMoV, however, is not an ideal population biomarker, especially for small towns due to its

557 large temporal variation.

558 Many wastewater SARS-CoV-2 surveillance studies have been conducted across the U.S. during the last two years³⁵, which provided an excellent network for the effective and long-term 559 560 monitoring of SARS-CoV-2 and possibly other diseases in the future. Thus far, however, the normalization applied to the SARS-CoV-2 wastewater surveillance does not have a standardized 561 562 approach. There are three main types of normalization: 1) normalized to WWTP flow (e.g., 563 copies/week, to give viral load), 2) normalized to WWTP human fecal biomarker estimated 564 population (e.g., copies/10K people/week), and 3) directly normalized to WWTP human fecal 565 biomarker loads (e.g., copies/copy of PMMoV/week).

566 Wastewater flow normalization converts the measured viral concentration to viral load, 567 accounting for variations in flow between days due to precipitation, snowmelt, or groundwater 568 inflow. Established on the flow normalization, the normalization to biomarker estimated 569 population evaluated in this study aims to account for the variations caused by wastewater 570 volume and population size that contribute to the waste. The third normalization is used often 571 because it does not involve wastewater volume and population information. However, it is based 572 on the assumption that the measured wastewater biomarker concentration perfectly reflects the 573 population dynamics. Many previous studies used PMMoV as SARS-CoV-2 internal control to normalize SARS-CoV-2 concentration to SARS-CoV-2 copies per copy of PMMoV.^{4,18,49} 574 575 However, our study showed that the relationship between PMMoV and population is not as 576 stable over time as caffeine and paraxanthine. In addition, our previous study confirmed that

577 direct normalization effects of SARS-CoV-2 concentration using biomarker concentrations were 578 always less ideal than indirect normalization, which involved the wastewater volume.²⁹

579 Our findings suggest that for long-term wastewater SARS-CoV-2 surveillance, normalizing

- 580 SARS-CoV-2 wastewater concentrations with a reliable population marker prior to calculating
- 581 trends is highly recommended to account for changes in wastewater dilution and differences in
- 582 relative human waste input over time due to tourism, weekday commuters, temporary workers,

It is made available under a CC-BY-ND 4.0 International license .

and pandemic lockdowns, etc. This approach is particularly critical for the sewershed with

584 dynamic population changes, such as colleges, tourist towns, and metropolitan areas, where a

585 large number of commuters who used to travel to cities daily transitioned to fully or partially

remote workers after the pandemic. The relation between 1) size of the population, and 2)

587 strength of the relationship between wastewater SARS-CoV-2 viral concentration and the

588 COVID-19 clinical incidence rate was first demonstrated in this study. This would provide an

589 excellent selection criterion for site selection, surveillance planning and data interpretation for

590 the SARS-CoV-2 and even other wastewater-based epidemiology.

591 **5. CONCLUSIONS**

592 This study compared the effectiveness of three wastewater biomarkers for population 593 normalization in the SARS-CoV-2 wastewater-based epidemiology with a large number of 594 wastewater facilities across Missouri and long-term sampling over seven months. We found that 595 PMMoV, one of the widely used population biomarkers for SARS-CoV-2 wastewater-based 596 epidemiology, is not an ideal population biomarker since it reduced the strength of wastewater 597 SARS-CoV-2 viral load and the COVID-19 incidence rate compared to the metadata population 598 and without population normalization. Instead, paraxanthine, with many benefits, such as high 599 stability, low exogenous sources than its precursor (caffeine), higher recovery rate, and low 600 quantification variation, is very promising for predicting the real-time population and population 601 normalization in the wastewater SARS-CoV-2 surveillance study, no matter the size of the 602 population. The utility of this promising biomarkers was validated by data from ten different 603 Wisconsin's WWTPs with gradients in population sizes. The estimated real-time population 604 using this biomarker was directly compared against the population patterns with human 605 movement mobility data. Of the three biomarkers, population normalization by paraxanthine 606 significantly strengthened the relationship between wastewater SARS-CoV-2 viral load and 607 COVID-19 incidence rate the most (40 out of 61 sewersheds). Our findings suggest that 608 paraxanthine has the potential to be serve as real-time population biomarker in other scenarios 609 beyond SARS-CoV-2 wastewater-based epidemiology and not limited within Missouri 610 geographical boundary.

611 6. ACKNOWLEDGEMENT

612 The authors would like to thank the Missouri Department of Health and Senior Services 613 (DHSS) for administrating the funding. We would like to express our gratitude to the Missouri 614 Department of Natural Resources (DNR) for coordinating the sample collection. Research 615 reported in this publication was supported by funding from the Centers for Disease Control and 616 the National Institute on Drug Abuse of the National Institutes of Health under award number 617 U01DA053893-01. We would also like to thank the Center for Agroforestry at the University of 618 Missouri, USDA/ARS Dale Bumpers Small Farm Research Center under agreement number 58-619 6020-6-001 from the USDA Agricultural Research Service for supporting part of this research. 620 The content is solely the responsibility of the authors and does not necessarily represent the 621 official views of the National Institutes of Health, the Centers for Disease Control or USDA-622 ARS.

It is made available under a CC-BY-ND 4.0 International license .

623 **7. REFERENCE**

- Choi PM, Tscharke BJ, Donner E, et al. Wastewater-based epidemiology biomarkers: Past,
 present and future. *TrAC Trends in Analytical Chemistry*. 2018;105:453-469.
 doi:10.1016/j.trac.2018.06.004
- Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the
 Early Stage of the Epidemic in The Netherlands. *Environ Sci Technol Lett.* 2020;7(7):511 516. doi:10.1021/acs.estlett.0c00357
- McMahan CS, Self S, Rennert L, et al. COVID-19 wastewater epidemiology: a model to
 estimate infected populations. *The Lancet Planetary Health*. 2021;5(12):e874-e881.
 doi:10.1016/S2542-5196(21)00230-8
- 4. Wu F, Zhang J, Xiao A, et al. SARS-CoV-2 Titers in Wastewater Are Higher than Expected
 from Clinically Confirmed Cases. 2020;5(4):9.
- Kumar M, Jiang G, Kumar Thakur A, et al. Lead time of early warning by wastewater
 surveillance for COVID-19: Geographical variations and impacting factors. *Chemical Engineering Journal*. 2022;441:135936. doi:10.1016/j.cej.2022.135936
- 6. Kirby AE, Walters MS, Jennings WC, et al. Using Wastewater Surveillance Data to
 640 Support the COVID-19 Response United States, 2020–2021. *MMWR Morb Mortal Wkly*641 *Rep.* 2021;70(36):1242-1244. doi:10.15585/mmwr.mm7036a2
- 642 7. Chan AY, Kim H, Bell ML. Higher incidence of novel coronavirus (COVID-19) cases in
 643 areas with combined sewer systems, heavy precipitation, and high percentages of
 644 impervious surfaces. *Science of The Total Environment*. 2022;820:153227.
 645 doi:10.1016/j.scitotenv.2022.153227
- 646 8. Gracia-Lor E, Castiglioni S, Bade R, et al. Measuring biomarkers in wastewater as a new
 647 source of epidemiological information: Current state and future perspectives. *Environment*648 *International.* 2017;99:131-150. doi:10.1016/j.envint.2016.12.016
- Medema G, Been F, Heijnen L, Petterson S. Implementation of environmental surveillance
 for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. *Current Opinion in Environmental Science & Health*. 2020;17:49-71.
 doi:10.1016/j.coesh.2020.09.006
- 10. Daughton CG. Real-time estimation of small-area populations with human biomarkers in
 sewage. *Science of The Total Environment*. 2012;414:6-21.
 doi:10.1016/j.scitotenv.2011.11.015
- Chen C, Kostakis C, Gerber JP, Tscharke BJ, Irvine RJ, White JM. Towards finding a
 population biomarker for wastewater epidemiology studies. *Science of The Total Environment*. 2014;487:621-628. doi:10.1016/j.scitotenv.2013.11.075

- McCall AK, Bade R, Kinyua J, et al. Critical review on the stability of illicit drugs in
 sewers and wastewater samples. *Water Research*. 2016;88:933-947.
 doi:10.1016/j.watres.2015.10.040
- Been F, Rossi L, Ort C, Rudaz S, Delémont O, Esseiva P. Population Normalization with
 Ammonium in Wastewater-Based Epidemiology: Application to Illicit Drug Monitoring.
 Environ Sci Technol. 2014;48(14):8162-8169. doi:10.1021/es5008388
- 14. Sweetapple C, Wade MJ, Grimsley JMS, Bunce JT, Melville-Shreeve P, Chen AS. *Population Normalisation in Wastewater-Based Epidemiology for Improved Understanding of SARS-CoV-2 Prevalence: A Multi-Site Study.* Epidemiology; 2021.
 doi:10.1101/2021.08.03.21261365
- 15. Udert KM, Larsen TA, Gujer W. Fate of major compounds in source-separated urine. *Water Science and Technology*. 2006;54(11-12):413-420. doi:10.2166/wst.2006.921
- kar 16. van Nuijs ALN, Mougel JF, Tarcomnicu I, et al. Sewage epidemiology A real-time
 approach to estimate the consumption of illicit drugs in Brussels, Belgium. *Environment International.* 2011;37(3):612-621. doi:10.1016/j.envint.2010.12.006
- 17. Holm RH, Nagarkar M, Yeager RA, et al. Surveillance of RNase P, PMMoV, and
 CrAssphage in wastewater as indicators of human fecal concentration across urban sewer
 neighborhoods, Kentucky. *FEMS Microbes*. 2022;3:xtac003. doi:10.1093/femsmc/xtac003
- b. D'Aoust PM, Mercier E, Montpetit D, et al. Quantitative analysis of SARS-CoV-2 RNA
 from wastewater solids in communities with low COVID-19 incidence and prevalence. *Water Research*. 2021;188:116560. doi:10.1016/j.watres.2020.116560
- Gasser G, Rona M, Voloshenko A, et al. Quantitative Evaluation of Tracers for
 Quantification of Wastewater Contamination of Potable Water Sources. *Environ Sci Technol.* 2010;44(10):3919-3925. doi:10.1021/es100604c
- 683 20. Oppenheimer J, Eaton A, Badruzzaman M, Haghani AW, Jacangelo JG. Occurrence and
 684 suitability of sucralose as an indicator compound of wastewater loading to surface waters in
 685 urbanized regions. *Water Research*. 2011;45(13):4019-4027.
 686 doi:10.1016/j.watres.2011.05.014
- Kitajima M, Iker BC, Pepper IL, Gerba CP. Relative abundance and treatment reduction of
 viruses during wastewater treatment processes Identification of potential viral indicators. *Science of The Total Environment*. 2014;488-489:290-296.
 doi:10.1016/j.scitotenv.2014.04.087
- Kitajima M, Sassi HP, Torrey JR. Pepper mild mottle virus as a water quality indicator. *npj Clean Water*. 2018;1(1):19. doi:10.1038/s41545-018-0019-5
- Baughton CG. Using biomarkers in sewage to monitor community-wide human health:
 Isoprostanes as conceptual prototype. *Science of The Total Environment*. 2012;424:16-38.
 doi:10.1016/j.scitotenv.2012.02.038

It is made available under a CC-BY-ND 4.0 International license .

696 697 698	24.	Thai PK, O'Brien JW, Banks APW, et al. Evaluating the in-sewer stability of three potential population biomarkers for application in wastewater-based epidemiology. <i>Science of The Total Environment</i> . 2019;671:248-253. doi:10.1016/j.scitotenv.2019.03.231
699 700 701	25.	Heckman MA, Weil J, de Mejia EG. Caffeine (1, 3, 7-trimethylxanthine) in Foods: A Comprehensive Review on Consumption, Functionality, Safety, and Regulatory Matters. <i>Journal of Food Science</i> . 2010;75(3):R77-R87. doi:10.1111/j.1750-3841.2010.01561.x
702 703	26.	Crews HM, Olivier L, Wilson LA. Urinary biomarkers for assessing dietary exposure to caffeine. <i>Food Addit Contam.</i> 2001;18(12):1075-1087. doi:10.1080/02652030110056630
704 705 706 707	27.	Martínez Bueno MJ, Uclés S, Hernando MD, Dávoli E, Fernández-Alba AR. Evaluation of selected ubiquitous contaminants in the aquatic environment and their transformation products. A pilot study of their removal from a sewage treatment plant. <i>Water Research</i> . 2011;45(6):2331-2341. doi:10.1016/j.watres.2011.01.011
708 709 710 711	28.	Senta I, Gracia-Lor E, Borsotti A, Zuccato E, Castiglioni S. Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. <i>Water Research</i> . 2015;74:23-33. doi:10.1016/j.watres.2015.02.002
712 713 714	29.	Hsu SY, Bayati MB, Li C, et al. <i>Biomarkers Selection for Population Normalization in SARS-CoV-2 Wastewater-Based Epidemiology</i> . Epidemiology; 2022. doi:10.1101/2022.03.14.22272359
715 716	30.	U.S. Census. Population Density in Missouri Counties: 2020. Published 2020. https://public.tableau.com/shared/CBTDRF3S8?:showVizHome=no
717 718	31.	Apple Maps. COVID-19 Apple Mobility Trends Reports. https://aws.amazon.com/marketplace/pp/prodview-ucdg4nudicxzi#offers
719 720 721	32.	Robinson CA, Hsieh HY, Hsu SY, et al. Defining biological and biophysical properties of SARS-CoV-2 genetic material in wastewater. <i>Science of The Total Environment</i> . 2022;807:150786. doi:10.1016/j.scitotenv.2021.150786
722 723	33.	R Core Team. <i>R: A Language and Environment for Statistical Computing</i> . R Foundation for Statistical Computing; 2021. https://www.R-project.org/.
724 725 726	34.	Ahmed W, Bivins A, Metcalfe S, et al. Evaluation of process limit of detection and quantification variation of SARS-CoV-2 RT-qPCR and RT-dPCR assays for wastewater surveillance. <i>Water Research</i> . 2022;213:118132. doi:10.1016/j.watres.2022.118132
727 728 729	35.	Duvallet C, Wu F, McElroy KA, et al. Nationwide Trends in COVID-19 Cases and SARS-CoV-2 RNA Wastewater Concentrations in the United States. <i>ACS EST Water</i> . Published online May 3, 2022:acsestwater.1c00434. doi:10.1021/acsestwater.1c00434
730 731	36.	Feng S, Roguet A, McClary-Gutierrez JS, et al. Evaluation of Sampling, Analysis, and Normalization Methods for SARS-CoV-2 Concentrations in Wastewater to Assess COVID-

It is made available under a CC-BY-ND 4.0 International license .

732 733		19 Burdens in Wisconsin Communities. ACS EST Water. 2021;1(8):1955-1965. doi:10.1021/acsestwater.1c00160
734 735 736 737	37.	Zhan Q, Babler KM, Sharkey ME, et al. Relationships between SARS-CoV-2 in Wastewater and COVID-19 Clinical Cases and Hospitalizations, with and without Normalization against Indicators of Human Waste. <i>ACS EST Water</i> . Published online May 26, 2022:acsestwater.2c00045. doi:10.1021/acsestwater.2c00045
738 739 740	38.	Rachmadi AT, Kitajima M, Pepper IL, Gerba CP. Enteric and indicator virus removal by surface flow wetlands. <i>Science of The Total Environment</i> . 2016;542:976-982. doi:10.1016/j.scitotenv.2015.11.001
741 742 743	39.	Buerge IJ, Poiger T, Müller MD, Buser HR. Caffeine, an Anthropogenic Marker for Wastewater Contamination of Surface Waters. <i>Environ Sci Technol</i> . 2003;37(4):691-700. doi:10.1021/es020125z
744 745 746	40.	Gracia-Lor E, Rousis NI, Zuccato E, et al. Estimation of caffeine intake from analysis of caffeine metabolites in wastewater. <i>Science of The Total Environment</i> . 2017;609:1582-1588. doi:10.1016/j.scitotenv.2017.07.258
747 748	41.	Senchina DS, Hallam JE, Kohut ML, Nguyen NA. Alkaloids and athlete immune function: Caffeine, theophylline, gingerol, ephedrine, and their congeners. Published online 2014:26.
749 750 751	42.	Colson P, Richet H, Desnues C, et al. Pepper Mild Mottle Virus, a Plant Virus Associated with Specific Immune Responses, Fever, Abdominal Pains, and Pruritus in Humans. Mylonakis E, ed. <i>PLoS ONE</i> . 2010;5(4):e10041. doi:10.1371/journal.pone.0010041
752 753 754	43.	Haramoto E, Kitajima M, Kishida N, et al. Occurrence of Pepper Mild Mottle Virus in Drinking Water Sources in Japan. <i>Appl Environ Microbiol</i> . 2013;79(23):7413-7418. doi:10.1128/AEM.02354-13
755 756 757	44.	Rosario K, Symonds EM, Sinigalliano C, Stewart J, Breitbart M. <i>Pepper Mild Mottle Virus</i> as an Indicator of Fecal Pollution. <i>Appl Environ Microbiol</i> . 2009;75(22):7261-7267. doi:10.1128/AEM.00410-09
758 759 760 761	45.	Kato R, Asami T, Utagawa E, Furumai H, Katayama H. Pepper mild mottle virus as a process indicator at drinking water treatment plants employing coagulation-sedimentation, rapid sand filtration, ozonation, and biological activated carbon treatments in Japan. <i>Water Research</i> . 2018;132:61-70. doi:10.1016/j.watres.2017.12.068
762 763 764 765	46.	Symonds EM, Sinigalliano C, Gidley M, Ahmed W, McQuaig-Ulrich SM, Breitbart M. Faecal pollution along the southeastern coast of Florida and insight into the use of pepper mild mottle virus as an indicator. <i>J of Applied Microbiology</i> . 2016;121(5):1469-1481. doi:10.1111/jam.13252
766 767	47.	Olesen S. Interpreting building-level Covid-19 wastewater monitoring data. Published online 2021:13.

It is made available under a CC-BY-ND 4.0 International license .

- 48. Bayati M, Hsieh HY, Hsu SY, et al. *Identification and Quantification of Bioactive Compounds Suppressing SARS-CoV-2 Signals in Wastewater-Based Epidemiology Surveillance*. Epidemiology; 2022. doi:10.1101/2022.03.09.22272155
- 49. Isaksson F, Lundy L, Hedström A, Székely AJ, Mohamed N. Evaluating the Use of
- 772 Alternative Normalization Approaches on SARS-CoV-2 Concentrations in Wastewater:
- Experiences from Two Catchments in Northern Sweden. *Environments*. 2022;9(3):39.
- 774 doi:10.3390/environments9030039

It is made available under a CC-BY-ND 4.0 International license .

No.	Compound	RT	ES	MS1	MS2	Cone Voltage	Collision Energy
1	Caffeine	6.273	ES+	195.05	138.12	45	22
2	Caffeine- ¹³ C ₃	6.167	ES+	198.04	140.07	45	22
3	Paraxanthine	5.715	ES+	181.06	124.11	45	22

776 **Table 1.** Summary of the optimized LC-MSMS Parameters for chemical population biomarkers.

It is made available under a CC-BY-ND 4.0 International license .

- Table 2. Summary of ratios between estimated real-time population by three biomarkers to the
- 779 Metadata population.

Ratios	Mean	Max	Min	SD
Caffeine-population/metadata population	1.248	79.669	0.008	1.862
Paraxanthine-population/metadata population	1.137	12.053	0.008	0.741
PMMoV-population/metadata population	1.471	75.203	0.00008	2.26

It is made available under a CC-BY-ND 4.0 International license .

- 781 **Table 3.** Coefficient of determinations (R^2) of the linear regression models between log-
- transformed wastewater SARS-CoV-2 RNA load (copies/week/10K person) and clinical
- 783 COVID-19 incidence rate (case number/week/10K person) within each WWTP without and with
- population normalization by metadata population, and the real-time populations estimated from
- 785 caffeine, paraxanthine, and PMMoV. The bold and underlined values are the highest R² for each
- 786 WWTP among five models. "N.S." indicates the not significant models. "N.A. indicated not
- 787 available information.

Facility ID	Popula	Total	Without	Metadata	Caffeine	Paraxanthine	PMMoV
	tion	cases	population	population	estimated	estimated	estimated
	Served		normalizatio	normalization	real-time	real-time	real-time
			n		population	population	population
					normalizatio	normalizatio	normalizatio
					n	n	n
ANON1	900	13	0.229	0.223	0.220	<u>0.257</u>	N.S.
ALBNY	1730	178	0.781	0.802	0.799	<u>0.827</u>	0.786
MEMPH	1822	24	0.206	0.197	0.214	0.232	0.196
MILAN	1960	83	0.470	<u>0.478</u>	0.401	0.473	0.287
WLOSP	2100	82	N.S.	N.S.	N.S.	N.S.	N.S.
TRYSE	2934	1015	0.761	0.683	0.569	<u>0.763</u>	0.741
WARSW	2976	127	0.136	0.160	0.215	0.254	0.242
CASVL	3300	155	0.399	0.368	0.337	<u>0.406</u>	0.280
CAROL	3784	285	0.539	0.542	0.593	<u>0.644</u>	0.561
CHARL	4000	102	N.S.	N.S.	N.S.	N.S.	N.S.
BROOK	4600	84	0.375	0.369	0.338	0.405	<u>0.556</u>
ELDON	4895	312	0.481	<u>0.496</u>	0.359	0.485	0.246
MACON	5471	95	0.297	0.309	0.351	0.452	N.S.
DEXTW	6000	211	0.620	<u>0.634</u>	0.467	0.576	0.340
ANON2	6155	429	0.462	<u>0.470</u>	0.420	0.445	0.136
PACIF	7001	260	0.675	0.676	0.603	0.685	<u>0.745</u>
SDLCN	7500	218	<u>0.728</u>	0.725	0.390	0.576	0.283
UNONW	7936	1208	0.738	0.740	0.764	<u>0.765</u>	0.697
MARSH	8000	539	0.505	0.503	0.471	<u>0.563</u>	0.524
NEVAD	8082	955	0.332	0.328	0.287	<u>0.376</u>	0.468
SDLNO	8250	80	N.S.	N.S.	N.S.	N.S.	N.S.
PRYVL	9000	645	0.391	0.393	0.328	<u>0.417</u>	0.322
KCTDC	9000	618	0.827	0.828	0.813	<u>0.851</u>	0.749
MONET	9100	338	0.693	0.693	0.687	<u>0.761</u>	0.634
MRSHL	10113	716	0.407	0.407	<u>0.567</u>	0.526	0.564
FRMTN	10114	775	0.590	0.590	0.544	<u>0.649</u>	0.549
BLIVR	10500	466	0.611	0.610	0.614	<u>0.645</u>	0.643
ANON3	10559	378	0.350	0.352	0.303	<u>0.417</u>	0.251
SDLSE	10730	225	0.7657	<u>0.7659</u>	0.464	0.697	0.306
MEXCO	11500	752	0.703	0.697	0.667	<u>0.707</u>	0.548
WARNE	11883	574	0.560	0.558	0.530	0.605	<u>0.832</u>
CARTH	12000	606	0.593	0.596	0.490	<u>0.646</u>	0.422
ANON4	12000	642	0.766	0.771	0.743	<u>0.791</u>	0.581
WPLAN	12000	726	0.879	0.882	0.882	0.893	<u>0.940</u>
KCROB	12000	665	0.610	0.616	0.591	<u>0.646</u>	0.410
KCFSR	12000	525	0.588	0.590	0.575	0.578	<u>0.639</u>
FULTN	12790	859	0.358	0.363	0.333	<u>0.402</u>	0.300
WARNW	14477	389	0.659	0.653	0.701	<u>0.721</u>	0.666
WASHN	15000	691	<u>0.659</u>	0.655	0.524	0.657	0.570
JOPSC	15906	426	0.729	0.733	0.737	<u>0.795</u>	0.675
HANBL	16000	727	0.414	0.416	0.490	<u>0.512</u>	0.307
SKSTN	17000	1289	0.674	0.680	<u>0.685</u>	0.677	0.570
NIXAF	20000	1118	0.680	0.678	0.669	<u>0.698</u>	0.693
SFDNW	22818	2826	0.731	0.727	0.715	0.733	<u>0.771</u>
MSDFN	24174	1633	0.515	0.527	0.435	<u>0.543</u>	0.376

It is made available under a CC-BY-ND 4.0 International license .

ROLSE 25423 1309 0.676 0.545 0.668 0.475 0.681 27000 0.787 0.787 0.740 0.533 STJOE 5048 0.851 NPSDS 27391 0.657 0.705 0.487 1841 0.656 0.613 CAPEG 33540 1653 0.591 0.599 0.474 <u>0.698</u> 0.401 JOPTC 34403 2176 0.872 0.861 0.852 0.871 0.766 0.8541 LIBTY 35300 1534 0.8537 0.853 0.811 0.749 **JEFFC** 41153 2861 0.759 0.766 0.776 0.806 0.645 0.800 STPSC 60000 3998 0.792 0.780 0.770 0.809 KCWST 61250 1926 0.826 0.814 0.798 0.840 0.792 0.795 MSDLM 0.789 0.790 0.726 0.728 66738 3677 0.883 KCBIR 76759 4455 <u>0.904</u> 0.902 0.798 0.777 MSDGG 115895 7392 0.797 0.787 0.809 0.764 0.664 COLMB 123180 6607 0.783 0.763 0.741 0.810 0.640 0.905 SFDSW 151966 8465 0.866 0.842 0.822 0.899 0.799 MSDMR 174537 11632 0.854 0.859 0.844 0.864 306647 0.756 **MSDBP** 15950 0.662 0.663 0.681 0.562 LBVAT 360000 16524 0.938 0.930 0.932 0.944 0.899 MSDME 451367 22557 0.874 0.872 <u>0.906</u> 0.888 0.723 **KCBLU** N.A. 8020 0.638 N.A. 0.510 <u>0.645</u> 0.409

It is made available under a CC-BY-ND 4.0 International license .

- Table 4. Case study in the smallest WWTP ANON1 (900 population) for the process limit of
- detection of wastewater SARS-CoV-2 surveillance. The process limit of detection (PLOD) of
- 791 N1N2 mass load PLOD1 for at least one case was calculated based on the results of this study
- (Figure 7B). The PLOD2 was calculated based on the literature.³⁴ The bold and underling values
- were the three weeks that N1N2 mass load higher both PLOD1 and PLOD2.

SWCTY	Flow rate				
Cases	(GMD)	Week	N1N2 mass load	N1N2 mass - PLOD1	N1N2 mass - PLOD2
0	0.0425	9/13/2021	5.56E+10	-7.44E+10	-3.35E+10
0	0.0639	9/20/2021	7.10E+10	-5.90E+10	-6.29E+10
0	0.0535	9/27/2021	9.22E+10	-3.78E+10	-1.99E+10
1	0.0531	10/04/2021	7.85E+10	-5.15E+10	-3.28E+10
0	0.0553	10/11/2021	1.62E+10	-1.14E+11	-9.97E+10
0	0.0488	10/18/2021	4.92E+10	-8.08E+10	-5.31E+10
0	0.0562	10/25/2021	1.38E+10	-1.16E+11	-1.04E+11
0	0.0546	11/8/2021	6.63E+09	-1.23E+11	-1.08E+11
0	0.0512	11/15/2021	1.45E+10	-1.16E+11	-9.28E+10
0	0.0474	11/22/2021	8.18E+10	-4.82E+10	-1.75E+10
1	0.0458	11/29/2021	7.91E+10	-5.09E+10	-1.69E+10
0	0.0493	12/6/2021	1.54E+10	-1.15E+11	-8.79E+10
2	0.053	12/13/2021	4.32E+09	-1.26E+11	-1.07E+11
<u>1</u>	<u>0.0544</u>	12/20/2021	<u>7.34E+11</u>	<u>6.04E+11</u>	<u>6.20E+11</u>
6	0.0544	1/10/2022	1.98E+12	1.85E+12	1.87E+12
$\overline{0}$	0.0573	1/31/2022	6.01E+10	-6.99E+10	-6.00E+10
<u>0</u>	<u>0.06</u>	<u>2/7/2022</u>	<u>1.37E+11</u>	<u>7.00E+09</u>	<u>1.13E+10</u>
$\overline{2}$	0.06	2/14/2022	9.48E+10	-3.52E+10	-3.09E+10
0	0.0608	2/21/2022	4.69E+09	-1.25E+11	-1.23E+11
0	0.06	2/28/2022	3.36E+08	-1.30E+11	-1.25E+11
0	0.072	3/7/2022	1.44E+09	-1.29E+11	-1.49E+11
0	0.07	3/14/2022	9.97E+08	-1.29E+11	-1.46E+11
0	0.11	3/21/2022	2.18E+09	-1.28E+11	-2.28E+11
0	0.05	3/28/2022	6.19E+08	-1.29E+11	-1.04E+11

It is made available under a CC-BY-ND 4.0 International license .



Fig. 1. A total of 2624 wastewater samples were collected from 64 wastewater treatment plants
 (WWTPs) across Missouri, USA.

It is made available under a CC-BY-ND 4.0 International license .





- study time from 09/13/2021 to 04/10/2022 (n=1569) within 64 WWTPs. The red lines indicate
- 801 the mean CV%, and the black lines indicate the median CV% of each biomarker.

It is made available under a CC-BY-ND 4.0 International license .





Fig. 3. The linear regression models between the mass loads of three wastewater biomarkers [(caffeine, paraxanthine, and pepper mild mottle virus (PMMoV)] and Metadata population

805 (GIS-mapped census population or estimated from sewer connections) for Missouri based on 7-

806 month wastewater samples collected from 09/13/21 to 04/10/22 across 63 wastewater treatment

807 plants (WWTPs) (plot A, C, and E). One facility from the initial 64 WWTPs (WWTP KCBLU)

808 was removed from the models since the population served changed dramatically during the

It is made available under a CC-BY-ND 4.0 International license .

- sampling period due to facility upgrading. Plots of B, D, and F were the models for Wisconsin
- 810 based on ten samples collected during the week of 06/07/2021 from ten different WWTPs.

It is made available under a CC-BY-ND 4.0 International license .



812 Fig. 4. Apple mobility indices of Columbia, MO for walking and driving using aggregated data

- 813 collected from Apple Maps data for the duration of 01/13/2020 to 04/10/2022 (A). The original
- 814 Apple mobility indices have been scaled to the maximum observed during the study period.
- 815 Predicted real-time population using paraxanthine in Columbia, MO (B).

It is made available under a CC-BY-ND 4.0 International license .



816

Fig. 5. The WWTP numbers with best population normalization models (highest R^2) for each

818 normalization model. The "61" after slash indicated the 61 out of 64 WWTPs had significant

819 linear regression models between wastewater SARS-CoV-2 RNA load (copies/week/10K

person) and clinical COVID-19 incidence rate (case/week/10K person). The numbers before the

slash indicated the number of WWTPs with the highest R^2 among the five regression models.



2

MILAN: 83/1960

WLOSP: 82/2100

TRYSE: 1015/2934

MEMPH: 24/1822

ALBNY: 178/1730

ANON1: 13/900

2

1

medRxiv preprint doi: https://doi.org/10.1101/2022.09.08.22279459; this version posted September 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.



- 823 Fig. 6. Normalized wastewater SARS-CoV-2 RNA load (copies/week/10K person) by
- 824 paraxanthine estimated real-time population and clinical COVID-19 incidence rate
- 825 (case/week/10K person) within 64 WWTPs from the weeks of 09/13/2021 to 04/05/2022. The
- 826 population of WWTPs increase from top left to right down. The title of each plot is in the format
- 827 of "WWTP name: total COVID-19 case number/Metadata population".

It is made available under a CC-BY-ND 4.0 International license .



828

Fig. 7. The polynomial relationships between the coefficient of determination (R^2) of the linear

830 regression models and clinical case (A) and Metadata population (B). The models were linear

831 regressions between wastewater SARS-CoV-2 RNA load (copies/week/10K person) and clinic

832 COVID-19 incidence rate (cases/week/10K person) within 64 WWTPs without and with the

833 population normalizations.

It is made available under a CC-BY-ND 4.0 International license .

Table S1. Long-time monitoring wastewater facilities across Missouri state and population sizes it served.

Facility ID	Facility Name	City	County	Population Served	Source of Population	Facility Capacity (MGD)	Composite sampling mode
ANON1	Anonymous WWTP#1			900	Operator information	0.14	Time Based
ALBNY	Albany WWTF	Albany	Gentry	1,730	Operator information	0.49	Time Based
MEMPH	Memphis Municipal WWTF	Memphis	Scotland	1,822	Operator information	0.21	Time Based
MILAN	Milan WWTP	Milan	Sullivan	1,960	Operator information	0.7	Flow Based
WLOSP	Willow Springs WWTP	Willow Springs	Howell	2,100	Operator information	0.4	Time Based
TRYSE	Troy Southeast WWTP	Troy	Lincoln	2,934	Connections w/ pop. Correction	0.45	Flow Based
WARSW	Warsaw WWTF	Warsaw	Benton	2,976	Operator information	1.1	Time Based
CASVL	Cassville WWTP	Cassville	Barry	3,300	Operator information	1.5	Time Based
CAROL	Carrollton WWTP	Carrollton	Carroll	3784	Operator information	1.5	Time Based
CHARL	Charleston	Charleston	Mississippi	4,000	Operator information	2	Time Based
BROOK	Brookfield WWTP	Brookfield	Linn	4,600	Operator information	1	Time Based
ELDON	Eldon WWTP	Eldon	Miller	4,895	Operator information	2.5	Time Based
MACON	Macon WWTP	Macon	Macon	5,471	Operator information	0.78	Time Based
DEXTW	Dexter West WWTP	Dexter	Stoddard	6,000	Operator information	3.4	Time Based
ANON2	Anonymous WWTP#2			6,155	Operator information	2	Time Based
PACIF	Pacific WWTP	Pacific	Franklin	7,001	Operator information	3.03	Time Based
SDLCN	Sedalia Central WWTP	Sedalia	Pettis	7,500	Connections w/ pop. Correction	1.5	Time Based
UNONW	Union West STP	Union	Franklin	7,936	Operator information	1.5	Time Based
MARSH	Marshfield WWTP	Marshfield	Webster	8,000	Connections w/ pop. Correction	2	Time Based
NEVAD	Nevada WWTF	Nevada	Vernon	8,082	Operator	2.5	Time Based
SDLNO	Sedalia North WWTP	Sedalia	Pettis	8,250	Operator	1.8	Time Based
PRYVL	Perryville SE WWTP	Perryville	Perry	9,000	operator	3.4	Time Based
KCTDC	KC, Todd Creek WWTP	Kansas City	Platte	9,000	information	6	Time Based
MONET	Monett Wastewater Treatment Plant	Monett	Barry	9,100	operator	1	Time Based
MRSHL	Marshall SE WWTP	Marshall	Saline	10,113	Operator	7.1	Time Based
FRMTN	Farmington East WWTP	Farmington	St. Francois	10,114	information	2.55	Time Based
BLIVR	Bolivar WWTP	Bolivar	Polk	10,500	information	5.3	Time Based
ANON3	Anonymous WWTP#3			10,559	pop. Correction	2.6	Time Based
SDLSE	Sedalia Southeast WWTP	Sedalia	Pettis	10,730	Operator information	3	Time Based
MEXCO	Mexico WWTP	Mexico	Audrain	11,500	information	1.5	Flow Based
WARNE	Warrensburg East WWTP	Warrensburg	Johnson	11,883	information	7	Time Based

It is made available under a CC-BY-ND 4.0 International license .

CARTH	Carthage WWTP	Carthage	Jasper	12,000	Operator	0	T. D. 1
	6	8	1	,	information	9	Time Based
ANON4	Anonymous WWTP#4			12,000	Operator	2	Time Deced
					Operator	3	Time Dased
WPLAN	West Plains WWTP	West Plains	Howell	12,000	information	2.8	Time Based
KCROB	KC Rocky Branch WWTP	Kansas City	Clay	12,000	Operator information	2	Time Based
KCFSR	KC, Fishing River WWTF	Kansas City	Clay	12,000	Operator information	2.9	Time Based
EUI TN	Fulton WWTD	Fulton	Collowov	12 700	Operator		
FULIN	Warrensburg West	Futton	Callaway	12,790	information Operator	1.5	Flow Based
WARNW	WWTP	Warrensburg	Johnson	14,477	information	4.8	Time Based
WASHN	Washington WWTP	Washington	Franklin	15,000	information	4	Time Based
JOPSC	Joplin Shoal Creek WWTP	Joplin	Newton	15,906	Connections w/ pop. Correction	7.2	Time Based
HANBL	Hannibal WWTP	Hannibal	Marion/Ralls	16,000	Operator	12	Time Based
SKSTN	Sikeston WWTP	Sikeston	Scott	17,000	Operator	2.75	T' D 1
				ŕ	Information	2.75	Time Based
NIXAF	Nixa WWTF	Nixa	Christian	20,000	information	5	Time Based
SFDNW	Springfield NW WWTP	Springfield	Greene	22,818	Operator	4	Time Deced
					Connections w/	4	Time Based
MSDFN	MSD Fenton WWTP	St. Louis	St. Louis	24,174	pop. Correction	6.8	Time Based
POLSE	Dollo SE WWTD	Pollo	Dhalma	25 422	Operator		
KOLSE	Kolla SE w w IF	Kolla	rneips	25,425	information	6.75	Time Based
STJOE	St. Joseph Water	St. Joseph	Buchanan	27,000	Connections w/	27	Time Deced
	NPSD Interim Saline	-			Connections w/	27	Time Dased
NPSDS	Creek Regional WWTP	FENTON	Jefferson	27,391	pop. Correction	4	Time Based
CAPEG	Cape Girardeau Municipal	Cape Girardeau	Cape Girardeau	33,540	Connections w/	15	Time Based
IODTO	Joplin Turkey Creek	J 1	J	24 402	Operator	15	Time Dased
JOPIC	WWTP	Joplin	Jasper	34,403	information	5	Time Based
LIBTY	Liberty WWTP	Liberty	Clay	35,300	Operator		
	5	Jefferson	5)	Information	11	Flow Based
JEFFC	Jefferson City RWRF	City	Callaway	41,153	information	9.5	Time Based
STPSC	St. Peters Spencer Creek	St. Peters	St. Charles	60,000	Connections w/	22.5	Time Deced
	wwiP				Operator	22.3	Time Dased
KCWST	KC, Westside WWTP	Kansas City	Jackson	61,250	information	15	Time Based
MSDI M	MSD Lower Meramec	St. Louis	St. Louis	66 738	Operator		
MODEM	WWTP	St. Louis	St. Louis	00,750	information	11	Flow Based
KCBIR	KC Birmingham WWTP	Kansas City	Clay	76,759	Connections w/	20	Time Deced
	MSD Grand Glaize				Operator	20	Thic Based
MSDGG	WWTP	Valley Park	St. Louis	115,895	information	21	Time Based
COI MB	Columbia WWTP	Columbia	Boone	123 180	Operator		
COLMB		Columbia	Doone	125,100	information	20.6	Time Based
SFDSW	Springfield SW WWTP	Springfield	Greene	151,966	pop. Correction	64	Time Based
MSDMR	MSD Missouri River WWTP	St. Louis	St. Louis	174,537	Operator	38	Time Based
MCDDD	MSD Bissell Point	St Lauia	St. Louis	206 617	Operator	50	Babea
MSDBP	WWTP	St. Louis	City	300,047	information	150	Time Based
LBVAT	LBVSD Atherton WWTP	Independence	Jackson	360,000	Population	50	Time D 1
		-			Operator	52	1 ime Based
MSDME	MSD Lemay WWTP	St. Louis	St. Louis	451,367	information	210	Time Based
KCBLU	KC Blue River WWTP	Kansas City	Jackson	490,000	Operator	105	T' D 1
		5			information	105	11me Based

It is made available under a CC-BY-ND 4.0 International license .

- 837 Table S2. Comparison of caffeine and paraxanthine concentrations from two sample preparation
- 838 methods with samples of weeks of December 6, 2021. "M" indicated dilution solution was
- methanol, and the corresponding sample above with the same sample I.D. and date (e.g., 1-12/7)
- 840 but without "M" means method with acidification and dilution solution was LC-MSMS buffer.

Samples lab I.D.	Caffeine concentration (ug/L)	Paraxanthine concentration (ug/L)
1-12/7	53.7	7 12.47
1-12/7M	54.3	7 9.54
2-12/6	130.5	7 20.00
2 12/6M	128.0	17 16
2 12/01	142.0	21.28
3-12/7	143.9	21.28
3-12/7M	158.4	22.45
4-12/7	59.6	4 9.66
4-12/7M	66.0	7 8.59
5-12/6	108.2	5 14.51
5-12/6M	111.0	8 14.71
6-12/7	42.4	1 8.79
6-12/7M	45.3	5 12.72
7-12/8	22.4	8 0.66
7 12/8M	24.5	0.00
0.12/6	24.3	0.52
9-12/6	63.9	+ /./1
9-12/6M	63.9	3 7.32
10-12/7	84.4	2 8.80
10-12/7M	84.4	9.60
11-12/6	60.3	8 16.19
11-12/6M	61.4	9 18.10
12-12/8	48 5	3 4 08
12-12/8M	40.5	4 5 5 1
14 12/7	4/.1	J.JI
14-12/7	115.8	25.76
14-12//M	121.4	23.36
15-12/7	70.2	5 5.61
15-12/7M	77.8	7 6.26
16-12/7	122.9	3 22.27
16-12/7M	135 1	4 22.38
18-12/7	87.0	1971
18-12/7 18-12/7M	08.7	24.42
10-12/7	98.7	24.42
19-12/7	/8.8	5 15.28
19-12/7M	83.2) 16.23
20-12/7	98.7	13.56
20-12/7M	105.3	2 13.75
22-12/6	101.5	1 15.80
22-12/6M	107.0	8 18.08
23-12/6	51.8	1990
23 12/6M	57.0	1 14.51
23-12/01/1	37.0	1 14.51
24-12/6	27.9	8.05
24-12/6M	29.0	7.06
25-12/7	107.6	8 21.25
25-12/7M	112.4	9 25.92
26-12/7	78.3	7 14.44
26-12/7M	84.8) 15 19
27-12/7	141 1	29.72
27 12/7M	146.0	1 29.72
27-12/7WI	140.9	20.04
28-12/7	04.8	/ /.8/
28-12//M	70.2	+ 7.71
29-12/7	45.0	3 10.61
29-12/7M	48.5	5 12.92
30-12/8	74.6	5 11.97
30-12/8M	75.5	1 15.27
31-12/7	70.3	5 10.81
31-12/7M	70.0	5 10.01
2/ 12/6	/ 7.7	- IU.I.5
34-12/0 24.12/CM	86.2	13.51
54-12/0M	91.1	13.74
35-12/7	52.3	3 8.81
35-12/7M	60.8	9.27
36-12/7	58.3	9 5.94
36-12/7M	66.1	1 6.94
37-12/6	72.3	3 613
37-12/6M	72.5	1 7.06
38 12/7	/1.0) 16.20
30-12/7	10/3.5	16.30
38-12//M	1164.2	16.38
39-12/7	36.3	9 0.00
39-12/7M	39.2	4 0.00
41-12/8	71.5	7 10.40
41-12/8M	73 3	7 10.75
42-12/7	493	9 8 23
12 12/7M	49.5 56 0	0.23
72-12//IVI	56.0	2 8.30

It is made available under a CC-BY-ND 4.0 International license .





It is made available under a CC-BY-ND 4.0 International license .



Fig. S2. A: Predicted real-time populations in the area served by the Kansas City Blue River
wastewater treatment plant (WWTP KCBLU) using concentrations of wastewater biomarkers
based on the models with the data from WWTP KCBLU. The metadata population was plotted
using the orange line for reference. The total Covid-19 clinical cases of each WWTP from
09/13/2021 to 04/10/2022 were strongly increased with the metadata population size (B: with
WWTP KCBLU; C: without WWTP KCBLU).

It is made available under a CC-BY-ND 4.0 International license .



- 856
 857 Fig. S3. Comparison of the coefficient of determination of the linear regression models (R²)
- between wastewater SARS-CoV-2 RNA load (copies/week/10K person) and clinic COVID-19
 incidence rate (cases/week/10K person) within 64 WWTPs without and with the population
- 860 normalizations

It is made available under a CC-BY-ND 4.0 International license .







Fig. S4. The linear regression relationship between wastewater SARS-CoV-2 RNA load with

- clinical cases (A: without population normalization; B: paraxanthine population normalization).
- The slopes of the regression equation indicated wastewater SARS-CoV-2 RNA concentration per week and person.