Evaluating Health Equity in Sub-City Wastewater Monitoring of COVID-19 in Davis, California, Through an Assessment of Demographics

By

AMITA MURALIDHARAN

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Heather N. Bischel, Chair

_____

Colleen E. Bronner

_____

Frank J. Loge

Committee in Charge

2024

ACKNOWLEDGMENTS

This research was conducted on land that has been the home of Patwin people for thousands of years. Today, there are three federally recognized Patwin tribes: Cachil DeHe Band of Wintun Indians of the Colusa Indian Community, Kletsel Dehe Wintun Nation, and Yocha Dehe Wintun Nation. We are honored and grateful for the stewardship of this land by the Patwin people over many centuries. Wastewater-based disease surveillance in Davis, CA and associated research was funded through Healthy Davis Together (HDT) and led by Professor Heather Bischel at the University of California, Davis. We greatly appreciate the dedication of the many City of Davis staff who were critical to the implementation of wastewater surveillance in Davis during the pandemic.

I thank Professors Heather Bischel, Frank Loge, and Colleen Bronner for serving on my thesis committee. I also thank Rachel Olson, Winston Bess, Hannah Safford, Lezlie Rueda, and the rest of the current and former Bischel Lab group members for their diverse contributions to this body of work.

ABSTRACT

Monitoring wastewater for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—whether performed at a building, neighborhood, or city level—has emerged as a viable way to track the prevalence of Coronavirus Disease-2019 (COVID-19) in a population. This study assesses health equity implications for wastewater sampling paradigms at a sub-city (or sub-*sewershed*) scale. Many wastewater-based disease surveillance efforts established during the COVID-19 pandemic relied on convenient points of access to sampling locations within the sewer network and/or voluntary participation within a city or region. Sampling in this way may generate public health data that does not equitably represent diverse populations within the area of interest. In preparation for future pandemics, better strategies are needed to design wastewater sampling frameworks. We help address this knowledge gap by: (1) developing a geospatial analysis tool that probabilistically assigns demographic data for subgroup populations aggregated by race and age (which were cited as major risk factors for severe COVID-19 outcomes) from census blocks to sub-sewershed sampling zones; (2) evaluating the representativeness of subgroup populations and sub-sewershed wastewater data in Davis, California, within the sampling framework employed for COVID-19 disease surveillance; and (3) demonstrating a scenario planning strategy in which adaptive sampling prioritizes vulnerable populations (in this case, populations >65 years old). Extensive sub-sewershed wastewater monitoring data was collected in Davis from November 2021 through September 2022, with wastewater samples collected three times per week from 15 maintenance holes (nodes) and daily from the influent of the city's centralized wastewater treatment plant (WWTP). The sub-city scale sampling achieved near complete coverage of the population, with spatial resolution

that informed public health communication initiatives within the city. Wastewater data aggregated from the sub-city scale as a population-weighted mean correlated strongly with wastewater data collected from the centralized treatment plant (Spearman's Rank correlation coefficient 0.909). We considered four down-scaling scenarios for a reduction in the number of sampling zones from baseline by 25% and 50%, chosen either randomly or by prioritizing maintenance of coverage of the >65-year-old population. Prioritizing representation of this vulnerable population in zone selection increased coverage of >65-year-olds from 51.1% to 67.2% when removing half of the sampling zones, while simultaneously increasing coverage of Black or African American populations from 67.5% to 76.7%. Downscaling the number of sampling zones had little effect overall on the correlation between the sub-sewershed zone wastewater data and centralized WWTP data (Spearman's Rank correlations ranged from 0.875 to 0.917), but the strongest correlations were obtained when maintaining sampling zones to maximize coverage of the >65-year-old population. When resource constraints necessitate downscaling the number of sampling sites, the approach demonstrated herein can inform decisions in ways that help preserve spatial representation of vulnerable populations, thereby promoting more inclusive, region-specific, and sustainable wastewater monitoring in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APD	Absolute percent difference
BCoV	Bovine Coronavirus
BSL2	Biosafety level 2
CDC	Centers for Disease Control and Prevention
CDPH	California Department of Public Health
COD	City of Davis
COVID-19	Coronavirus Disease-2019
ddPCR	Digital droplet polymerase chain reaction
FAM	Carboxyfluorescein
GEOID	Geographic identifier
HDT	Healthy Davis Together
HEX	Hexachlorofluorescein
LOB	Limit of blank
LOD	Limit of detection
MAE	Mean absolute error
MH	Maintenance hole
N	Nucleocapsid
PMMoV	Pepper mild mottle virus
PWMA	Population-weighted moving average
RNA	Ribonucleic acid
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SDOH	Social determinants of health
SR	Sub-city sewershed zone (or sub-sewershed zone)
USCB	United States Census Bureau
VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein
WBE	Wastewater-based epidemiology
WWTP	Wastewater treatment plant

INTRODUCTION

Social determinants of health (SDOH) have been identified as the set of conditions that drive health disparities (Holm, Osborne Jelks, et al., 2023). SDOH include a variety of factors such as race, socioeconomic position, and age, all of which play a crucial role in determining health outcomes within a population (Heller et al., 2014). The cumulative impact of these factors drives structural inequities (Mandelbaum, 2020). The Coronavirus Disease-2019 (COVID-19) pandemic highlighted how social vulnerabilities can contribute to infections and subsequent community spread, with increased risk of serious illness attributed to pre-existing comorbidities and older age, and disproportionate impacts seen in low-income populations and communities of color (Holm, Pocock, et al., 2023).

According to the Centers for Disease Control and Prevention (CDC), the three major factors affecting COVID-19's unequal distribution of impact are age, race, and ethnicity, with age cited as the main risk factor for severe COVID-19 outcomes (CDC, 2020). Data from the National Vital Statistics System has shown that, relative to those between ages 18-29 years old, risk of death from COVID-19 is 25 times higher for people between ages 50-64 years, 60 times higher for people between 65-74 years, 140 times higher for people between 75-84 years, and 340 times higher for those 85 years and older (*COVID-19 Death Data and Resources - National Vital Statistics System*, 2023). Moreover, the COVID-19 pandemic underscored racial and ethnic health inequities. Individuals from racial and ethnic minority groups were disproportionately affected by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission, leading to increased rates of hospitalization, emergency room visits, and premature death compared to non-Hispanic White individuals (*COVID-19 Death Data and Resources - National*

Vital Statistics System, 2023). Since March 2020, the average daily increase in COVID-19 mortality was found to be much higher in rural U.S. counties amongst predominantly Black and Hispanic populations (Cheng et al., 2020). Among rural counties, those in the top quartile of percent Black populations had an average daily increase in COVID-19 mortality rates 70% higher than counties in the bottom quartile, and counties in the top quartile of percent Hispanic populations had an average daily increase that was 50% higher (Cheng et al., 2020).

Wastewater-based epidemiology (WBE), which involves analyzing community-pooled wastewater samples from centralized wastewater treatment plants or sewer collection systems for disease biomarkers, has emerged as a viable way to provide insight into population-level disease trends. WBE has been favored as a minimally invasive, anonymous, and cost-effective way to track virus spread compared to testing individuals within the population since approximately more than 80% of U.S. residents are on a piped sewer system (Holm, Osborne Jelks, et al., 2023; National Academies of Sciences, Engineering, and Medicine, 2023). Each wastewater sample can represent hundreds to over a million people depending on the sample collection location. Those infected with SARS-CoV-2, including asymptomatic, pre-symptomatic, and symptomatic individuals, can shed viral particles and associated ribonucleic acid (RNA) through fecal matter. SARS-CoV-2 RNA remains readily detectable in wastewater even though fecal-oral transmission has not been reported for this virus. WBE has thus filled gaps associated with underreporting of cases, and can serve as an early indicator of potential outbreaks (Kadonsky et al., 2023). Moreover, WBE has proven to be a more comprehensive approach to tracking viral outbreaks and community infections since it does not rely on community members having access to clinical testing services or seeking healthcare when they are

experiencing symptoms (CDC, 2023a). WBE has proven especially useful in resource-limited settings (e.g., where clinical testing services are constrained).

Trends in wastewater data at a sub-city level (e.g., at the census block level, the most granular level at which public demographic data can be obtained) have also been used in WBE to inform public health responses (Bureau, n.d.-d). When the pandemic first began, many WBE programs were established rapidly via academic and government partnerships, and sampling locations were not always selected in a methodical way. While the utility of WBE is evident, sampling paradigms that rely on convenient points of access within the sewer network may not equitably serve the public or public health, as some populations may be underrepresented depending on where sampling occurs. In fact, there may be similar disparities in monitoring efforts as there are disparities in access to clinical testing and vaccinations (Medina et al., 2022). Recent efforts in the field have acknowledged the importance of taking steps targeted at reducing inequities (*National Academies of Sciences, Engineering, and Medicine,* 2023), but standardized measures for assessing performance of WBE towards improving health equity are lacking.

Equitable protection of public health can be guided through inclusive wastewater surveillance efforts (Holm, Pocock, et al., 2023). Specific considerations are needed to promote inclusion of underrepresented groups and equity in responses to public health threats. Previous studies have noted a lack of evaluations regarding the extent to which vulnerable populations are included in wastewater monitoring programs (Hu et al., 2023). Ultimately, the success of any WBE program relies on the assurance that there is equitable representation of underserved communities in the sampling design. Evaluating demographic representation of sub-sewershed

zone populations can prove especially difficult because data at a census block level does not align with sewer networks. In other words, flows of wastewater from populations within census blocks depend on city-wide sewer system connections and do not conform to the zones of population that are represented from wastewater samples collected from maintenance holes (MHs) within a city.

This project offers a strategy for WBE sub-city (or sub-*sewershed*) health equity evaluations using Census data at a block-level with the goal of enhancing inclusivity in the design of sampling frameworks within the constraints of a sewer system. First, we developed a probabilistic assignment approach to determine the expected subgroup population that is represented by collecting samples at different locations in a city sewer system. We used sub-sewershed wastewater surveillance during the COVID-19 pandemic and demographic data in Davis, California (population ~66,850), to demonstrate the approach (*U.S. Census Bureau QuickFacts*, n.d.). Second, we assessed how trends in wastewater data at the sub-sewershed level in Davis compared to trends for the city as a whole, since sampling frameworks may seek to achieve representativeness of overall community disease trends in addition to achieving representation based on demographic characteristics. Finally, we evaluated scenarios in which adaptive sampling strategies are implemented to prioritize representation of high-risk or vulnerable populations under resource-constrained conditions. The overall framework offers a strategy to evaluate sub-city sampling designs for wastewater surveillance to enhance health equity goals.

MATERIALS AND METHODS

STUDY SETTING AND DESIGN

This study includes a retrospective evaluation of SARS-CoV-2 wastewater concentration data collected during the COVID-19 pandemic as part of the Healthy Davis Together (HDT) program in Davis, a city with a population of approximately 66,850 in Yolo County, California (*U.S. Census Bureau QuickFacts*, n.d.). HDT offered wide-spread, free saliva-based asymptomatic and symptomatic testing, conducting over 1.6 million COVID-19 tests for the community and the university across 120 locations including school-based sites and mobile clinics (Pollock et al., 2022). From September 2020 to September 2022, wastewater surveillance was conducted throughout the City of Davis (COD) at the city, sub-city, and building/neighborhood levels. Table 1 provides definitions for key terms associated with the sampling framework. At the city level, samples were collected from the influent to the COD Wastewater Treatment Plant (COD WWTP). At the sub-sewershed level, samples were collected from up to sixteen nodes that each represent a sub-sewershed within the COD. At the building/neighborhood level, samples were collected from up to seven additional nodes for building complexes or neighborhoods identified as priority areas by HDT and local officials for potential communication and/or health interventions. The number of sampling locations and frequency of sampling increased through time. By April 2021, HDT sampled daily from the COD WWTP and three times per week from MHs in each of the sub-sewershed and building/neighborhood zones. Safford et al. (2022) describe HDT wastewater surveillance conducted from September 2020 to June 2021, for which wastewater samples were collected at seven building/neighborhood locations, 16 sub-sewershed nodes, and the city level. These

samples were analyzed using quantitative polymerase chain reaction (qPCR). Daza-Torres et al. (2023) report city-level (COD WWTP) wastewater surveillance data from December 1, 2021, to March 31, 2022, and analyzed using droplet digital PCR (ddPCR). This study describes and analyzes HDT wastewater surveillance conducted from November 22, 2021, to September 30, 2022, for which wastewater samples were collected daily from the COD WWTP and three times per week at 15 sub-sewershed nodes. All samples in this study were analyzed using ddPCR as reported by Daza-Torres et al. (2023) and described below.

Table 1. Introductory terms and definitions.

Key Term	Definition
Sewershed	The area that contributes wastewater to a common end point. In this study, the sewershed refers to the area whose sewers flow to the City of Davis Wastewater Treatment Plant (COD WWTP).
Sewershed node	A maintenance hole (MH) that serves as a wastewater sampling location.
Sub-sewershed zone	The area or population represented by one or more sewershed nodes. Also referred to as sub-city sewershed zones.
Subgroup	A subset of the overall city population that shares a specific demographic characteristic (e.g., race or age).

SAMPLE COLLECTION

Sample collection was previously described by Safford et al. (2022). Samples at each sub-sewershed node were collected using insulated Hach AS950 Portable Compact Samplers (Thermo Fisher Scientific, USA), and were programmed to collect 30 mL of sample every 15 minutes. Samples from the COD WWTP were collected using Teledyne ISCO 5800 refrigerated autosamplers. Each sampler was programmed to collect 400 mL of wastewater every 15 "pulses." Each pulse was set at 10,000 gallons (Daza-Torres et al., 2023). Based on the influent

flow of 3.6 million gallons per day, about 24 pulses were expected per day. Each sample date recorded corresponds to the date that an autosampler program was completed. The COD WWTP provided 12 mL samples in new 15-mL polypropylene centrifuge tubes. The samples were stored at 4℃. Samples were transported to the analytical laboratory at UC Davis in coolers on ice and generally processed the same day. Samples were pasteurized for 30 minutes at 60℃ to mitigate biohazard risk while maintaining RNA quality. The samples were returned to 4℃ prior to processing, and then concentrated and extracted in a biosafety level 2 (BSL2)-certified laboratory.

SAMPLE PROCESSING

The sample concentration and extraction protocols were previously described by Daza-Torres et al. (2023). Each starting sample volume of 4.875 mL was deposited into a separate well of a KingFisher 24 deep-well plate (Thermo Fisher). Nuclease-free water was included as a sample to act as a contamination control during the concentration and extraction process (Borchardt et al., 2021). Five µL of a stock of inactivated encapsulated whole vaccine-strain Bovine Coronavirus (BCoV, Bovilis® Coronavirus vaccine) was spiked into each well. The vaccine stock was determined by ddPCR quantification to contain approximately 1.3×10^8 gc/mL. The initial BCoV vaccine stock arrived lyophilized and was reconstituted in 20 mL of provided diluent, then divided into 1.5 mL aliquots and stored at -80℃ before use. Fifty µL of Nanotrap® Enhancement Reagent 1 (Ceres Nanosciences product ER1 SKU #10111-30) were added alongside the spike. Then, 75 µL Nanotrap® Magnetic Virus Particles (Ceres Nanosciences) were added to each sample and a KingFisher Apex robot (Thermo Scientific) was used to concentrate the viruses from the samples using a protocol provided by Ceres (*Ceres - Protocols*, n.d.) with

minor modifications described herein. Beads were eluted into 400 µL of lysis buffer per sample from the MagMAX Microbiome Ultra Nucleic Acid Isolation Kit (Thermo Fisher). Samples were transferred to 96 deep-well Kingfisher plates and extracted on a Kingfisher Apex according to kit instructions, ending at a final volume of 100 µL sample in MagMAX Elution Solution. Typically, extracts were stored on ice and analyzed on the same day, but extracts were stored at -80℃ for up to one week if same-day analysis was not possible to preserve RNA quality. In this study, we did not include the recorded concentration of SARS-CoV-2 in our final dataset if more than 14 days passed between sample collection and processing.

EXTRACT ANALYSIS BY DROPLET DIGITAL PCR

Digital droplet polymerase chain reaction (ddPCR) was used to analyze the extracts. The targets analyzed were the N1 and N2 targeting regions of the nucleocapsid (N) gene of SARS-CoV-2, and Bovine Coronavirus (BCoV) and pepper mild mottle virus (PMMoV) to normalize the SARS-CoV-2 results. The primer and probe sequences for each target are listed in Table 2. Separate duplex assays were used to quantify N1/N2 and BCoV/PMMoV. The fluorophores used for detection were Carboxyfluorescein (FAM) for N1 and PMMoV, 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC) for N2, and Hexachlorofluorescein (HEX) for BCoV. The sample for the PMMoV/BCoV duplex was diluted 40× before loading due to the high concentrations of PMMoV in wastewater. A QX ONE ddPCR System (Bio-Rad) was used to perform the ddPCR amplifications in 20 µL reactions. Each reaction contained 15 µL of master mix from the One-Step RT-ddPCR Advanced Kit for Probe (Bio-Rad) with primer concentrations at 900 nM of each primer and probe concentrations at 250 nM of each probe, plus a 5 µL of sample extract or control. The cycling conditions for RT-ddPCR are listed in Table 3.

Table 2. Primer and probe sequences for the specified targets.

Target	Oligonucleotide	Primer Sequence (5' → 3')	Fluorophore
SARS-CoV-2 N1	Forward	GAC CCC AAA ATC AGC GAA AT	FAM
	Reverse	TCT GGT TAC TGC CAG TTG AAT CTG	
	Probe	ACC CCG CAT TAC GTT TGG TGG ACC	
SARS-CoV-2 N2	Forward	TTA CAA ACA TTG GCC GCA AA	VIC
	Reverse	GCG CGA CAT TCC GAA GAA	
	Probe	ACA ATT TGC CCC CAG CGC TTC AG	
BCoV	Forward	CTGGAAGTTGGTGGAGTT	HEX
	Reverse	ATTATCGGCCTAACATACATC	
	Probe	CCTTCATATCTATACACATCAAGTTGTT	
PMMoV	Forward	GAGTGGTTTGACCTTAACGTTTGA	FAM
	Reverse	TTGTCGGTTGCAATGCAAGT	
	Probe	CCTA+C+C+GAAGCA+A+A+TG	

Table 3. Cycling conditions for the RT-ddPCR process.

Cycling step	Temp (℃)	Time	# Cycles
Plate equilibrium	25	3 min	1
Reverse Transcription	50	60 min	1
Enzyme activation	95	10 min	1
Denaturation	94	30 sec	40
Annealing/Extension	58	1 min	
Enzyme Deactivation	98	10 min	1
Droplet Stabilization	25	1 min	1

To avoid contamination, preparation and plating of the ddPCR master mix were conducted manually in a separate location from sample loading, which was performed using an epMotionR 5075 (Eppendorf) liquid handler. Each reaction plate included duplicate positive controls (stock mixture of synthesized gene fragments for the target regions) and duplicate no-template controls (nuclease-free water). Results were analyzed using the QX One Software Regular Edition (Bio-Rad) and thresholds were adjusted by visual inspection in samples and controls. The results were considered invalid if the distribution of positive or negative droplets appeared abnormal or if the total number of droplets generated was below 10,000 droplets in a well.

WASTEWATER DDPCR DATA PROCESSING

To assess the sensitivity of the assay, a limit of detection (LOD) and a limit of blank (LOB) were determined using protocols recommended by the ddPCR manufacturer (*Introduction to Digital PCR | Bio-Rad*, n.d.). Fifteen wastewater samples that were negative for SARS-CoV-2 were used to determine the lowest detectable SARS-CoV-2 concentrations in apparently negative wastewater samples. These 24-hour composite samples had been collected on November 11, 2021, from building clean-outs or MHs downstream of UC Davis residential buildings. The samples originally tested negative (zero positive droplets) or very low (no greater than three positive droplets for each N target). The samples were re-analyzed by ddPCR to acquire data for four additional replicates for each sample. Since the results from the blank were not normally distributed, rank order was used to select the LOB. To do so, the number of droplets from individual wells were recorded from lowest to highest, and the LOB was set at the value of the concentration measurement for the rank position corresponding to the 95th percentile, using the following equation:

$$Rank = 0.5 + 0.95 * (number\ of\ measurements) \qquad \text{(Equation 1)}$$

The rank position was rounded up to provide a more conservative LOB since the calculated value was a non-integer value. The theoretical LOD was determined by adding two times the standard deviation of all the replicate results to the LOB. The LOD and LOB values are reported in Table 4. The highest number of positive droplets in the merged wells of the blank samples were 6 (N1) and 8 (N2). The cutoff was set at 3 (N1) and 4 (N2) since wastewater samples were routinely analyzed in duplicate. This way, it was possible to mark samples below the droplet threshold. Additionally, if the samples had fewer N1 and N2 droplets than twice the number of

droplets in the extraction control blank that was analyzed on the same day, they were considered below the droplet threshold. Furthermore, runs that had an extraction control blank with greater than 15 positive droplets (N1 or N2) were considered contaminated, and the extracts were re-processed.

If the samples satisfied the criteria above, the relative concentration of N gene was calculated. This was done by merging the duplicate results for each target and calculating the concentration of each target in the ddPCR reaction, assuming a Poisson distribution using the QXOne Software 1.1.1 Standard Edition. To obtain the average SARS-CoV-2 RNA concentration in the initial wastewater sample, the N1 and N2 results were averaged after correcting for the sample and reagent volumes used. The resultant value was reported as genome copies (gc) per mL wastewater. Concentrations of targets were not corrected for BCoV recovery efficiency. A threshold BCoV recovery value of 10% was used to retain the recorded SARS-CoV-2 concentrations that had a recovery rate equal to or greater than the threshold value. High variability is to be expected among BCoV recovery values due to variability of sample characteristics. Other recovery analyses have reported average BCoV recovery values ranging from as low as 4.8% to 36.1%, depending on the virus concentration method employed (Juel et al., 2021). Targets were excluded from the average concentration if N1 or N2 merged droplet counts were below the minimum droplet threshold, and the concentration was reported as 0 if both the N1 and N2 targets were below the droplet threshold. We use N/PMMoV (the average SARS-CoV-2 RNA concentration (N) divided by the concentration of PMMoV) as the wastewater signal for subsequent analysis.

Table 4. Limit of blank (LOB, reported as a rank position value) and limit of detection (LOD), as reported in Daza-Torres et al. (2023).

		LoB	STD	Theoretical LOD
Conc (copies/mL of wastewater)	N1	13.605	5.547	24.700
	N2	18.544	6.712	31.967
Conc (copies/mL of wastewater)	N1	3.401	1.387	6.175
	N2	4.636	1.678	7.992

WASTEWATER DATASET

The dataset used in this study comprises normalized SARS-CoV-2 concentrations recorded between November 22, 2021, and September 30, 2022. The SARS-CoV-2 concentrations were normalized with PMMoV concentrations. PMMoV was universally present and in high concentrations in wastewater samples, making it a suitable biomarker for process control (Li et al., 2022). For visualization purposes, a 10-day trimmed right-aligned moving average was applied to the data. To do so, 10% of the values on either end of the dataset were removed prior to calculating the mean of the current day and the previous nine days. This effectively smoothed out short-term variations of the normalized N gene metric. Figure 1 depicts the wastewater data retrieved from one sub-sewershed zone (SR-A), showing the raw wastewater surveillance data (red) and the 10-day trimmed right-aligned moving average (black) of the data. In the figure, we can see that the peaks in the data coincide with the surge in infections when the Omicron BA.1 variant was predominant in the region between December 2021 and March 2022, and a similar surge between April 2022 and September 2022 when the BA.2, BA.4, and BA.5 subvariants were prevalent (CDC, 2023b). Similar patterns can be observed in the data available for the other sub-sewershed zones.

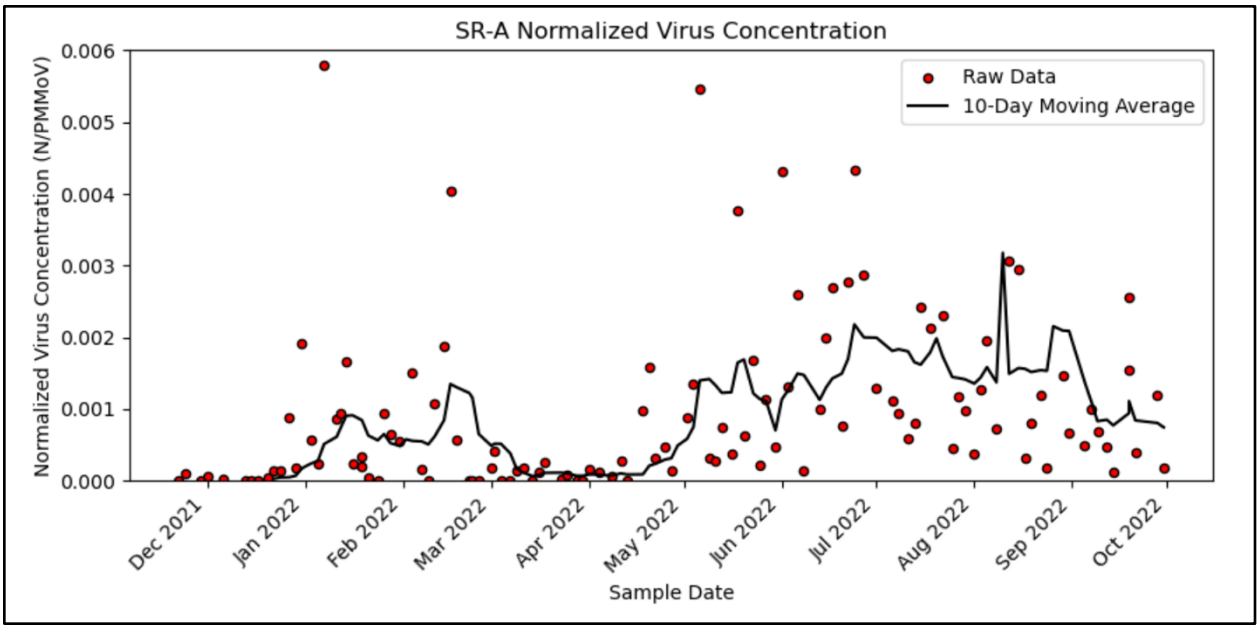


Figure 1. Normalized virus concentration for sub-sewershed zone A (SR-A).

SUB-SEWERSHED ZONES

The wastewater treatment plant area serviced by the COD WWTP was divided into sub-sewershed zones, which were delineated according to Figure 1 in Safford et al. (2022). While HDT conducted wastewater monitoring at both the sub-sewershed and building/neighborhood level, this study focuses on assessing trends across only the sub-sewershed zones. This is because most of the building/neighborhood zones are encompassed within the sub-sewershed zones, meaning that the building/neighborhood zones overlapped with the sub-sewershed zones. We aimed at selecting zones that spatially covered as much of the city as possible with minimal to no overlap of census block boundaries. Sub-sewershed zones were defined by a list of census block geographic identifiers (GEOIDs), which are numeric codes that uniquely identify each census block (Bureau, n.d.-c). Each zone definition comprises several GEOIDs. For the geospatial analysis conducted in this study, significant overlap of census blocks was observed for some sub-sewershed zones when defining them as they were delineated in Safford et al.

(2022). In these cases, the zones experiencing overlap were merged to form a larger sub-sewershed zone to prevent double counting populations for demographic analysis. This merging was performed by conglomerating the zone definitions for the smaller zones, and keeping all unique GEOIDs in the new, larger sub-sewershed zone definition. This is the case for SR-B (which merges SR-B1, SR-B2, SR-B3, and SR-B4), SR-C (merges SR-C1, SR-C2, and SR-C3), and SR-F (merges SR-F1 and SR-F2). The maintenance holes associated with each zone were then assigned to the overarching sub-sewershed zone from which the wastewater samples were being collected. The sub-sewershed zones and maintenance hole sampling sites are illustrated in Figure 2 below. These zones show how wastewater contributions from different parts of the city are isolated at the sub-city scale. The sampling locations (brown) can be seen atop the census blocks (lines). We then aligned the delineated zone boundaries with the respective census block boundaries across the city to be able to compare the demographics of the people residing in each zone.

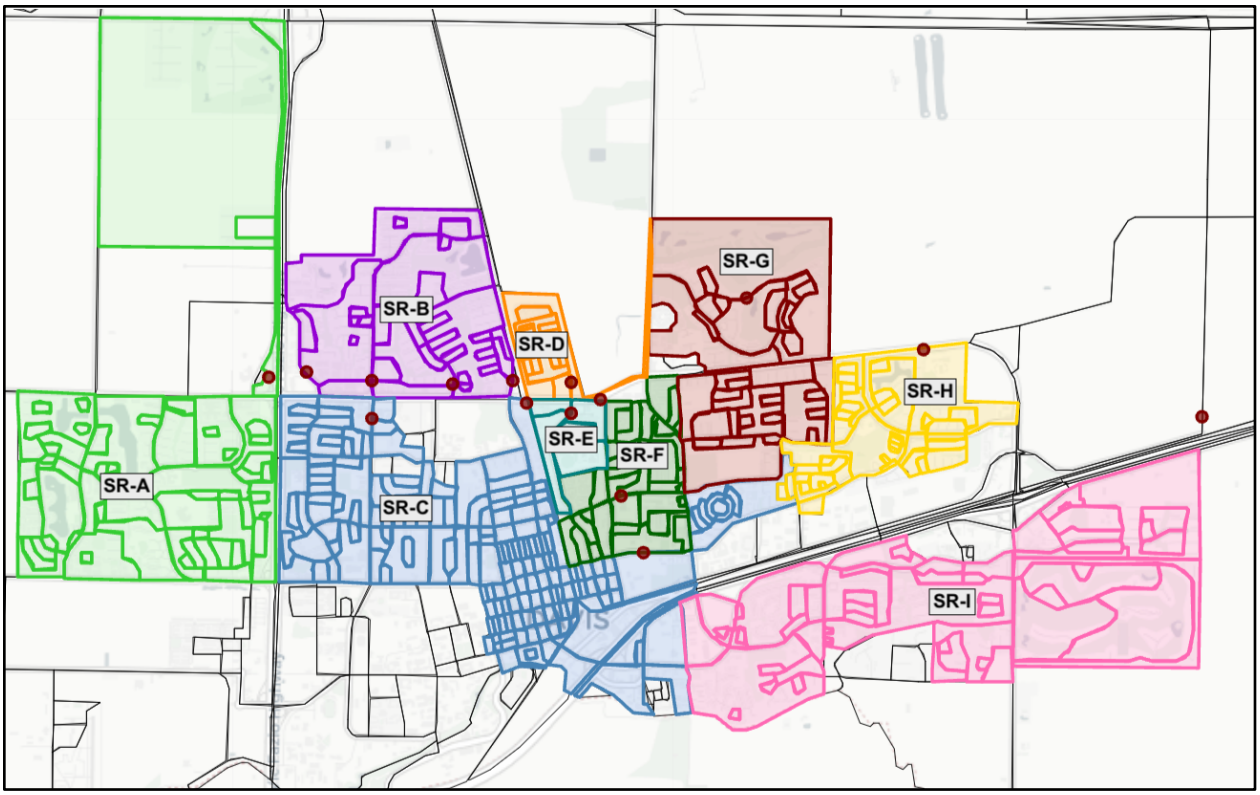


Figure 2. Map of the nine sub-sewershed zones for the SARS-CoV-2 wastewater-based epidemiology efforts in Davis (census blocks outlined).

DEMOGRAPHIC DATA

Demographic data was gathered from the U.S. Census Bureau (USCB), which includes census-block-level population data from 2020 (*Census Bureau Data*, n.d.). To prepare the Census data for analysis, we filtered the data to include only the census blocks within the COD, in addition to two areas within the catchment area that are designated as spheres of influence and thus should be included in our analysis (*Planning and Zoning | City of Davis, CA*, n.d.). We intersected the respective sub-sewershed zone areas with the census block boundaries. The total population within each sub-sewershed zone is listed in Table 5.

Table 5. Sub-sewershed zones and their populations.

Sub-sewershed Zone	Population
SR-A	12389
SR-B	7936
SR-C	15920
SR-D	1324
SR-E	1794
SR-F	3920
SR-G	5935
SR-H	3696
SR-I	10912

Because of the high spatial granularity of census-block-level data, it is important to note that only certain demographic factors had sufficient data available for this analysis. Additionally, increased margins of error were reported in the Census data collected in 2020 due to the difficulty of collecting responses during the COVID-19 pandemic (Bureau, n.d.-b). Consequently, this study focuses on the following two demographic factors: race and age. No other demographic variables were available in the 2020 Census data at the block level during the time of our analysis.

The tabular Census data was grouped according to the categories specified on the California Department of Public Health (CDPH) Health Equity Dashboard (*COVID-19 Age, Race and Ethnicity Data*, n.d.). The Census data presenting the racial composition of each census block was filtered to include the following seven groups: White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, Other, and Multi-Race. The Census data presenting the composition of people by age in each census

block was grouped into the following ten age categories: less than 5 years, 5-17 years, 18-34 years, 35-49 years, 50-59 years, 60-64 years, 65-69 years, 70-74 years, 75-79 years, and greater than or equal to 80 years.

All analyses were performed using Python version 3.11.5 (the Python script used for implementation is available at *https://tinyurl.com/HealthEquityWBE*).

PROBABILISTIC ASSIGNMENT TOOL

We adapted Python code from Safford et al. (2022) to combine information on municipal wastewater flows (provided as a File Geodatabase by the City of Davis Public Works Department) with Census data. The resultant probabilistic assignment tool assigns Census data to sewershed sampling zones based on geospatial probability. Each sampling zone spans several census blocks and may have different boundaries than the census blocks (e.g., in the case of overlap at neighboring zone boundaries). The probabilistic assignment tool performs the following tasks. First, the geospatial coordinates of all the MHs in the Davis sewer system and information about the relative upstream or downstream position of each MH were used to build a "connection graph" capturing directional connections among all the MHs. Second, the 2020 USCB population data was used to estimate the number of people living in each census block within the sampling zone of interest. Populations were spatially analyzed by age and race. We assumed that each person in each census block produces the same amount of wastewater every day, and that each person has an equal probability of discharging that wastewater to each MH within the block. We used the connection graph to probabilistically assign demographic data from census blocks to sewershed nodes. Thus, we extrapolated information

from the census block level to each sub-sewershed zone. See Figure 3 for an illustrative example of this approach.

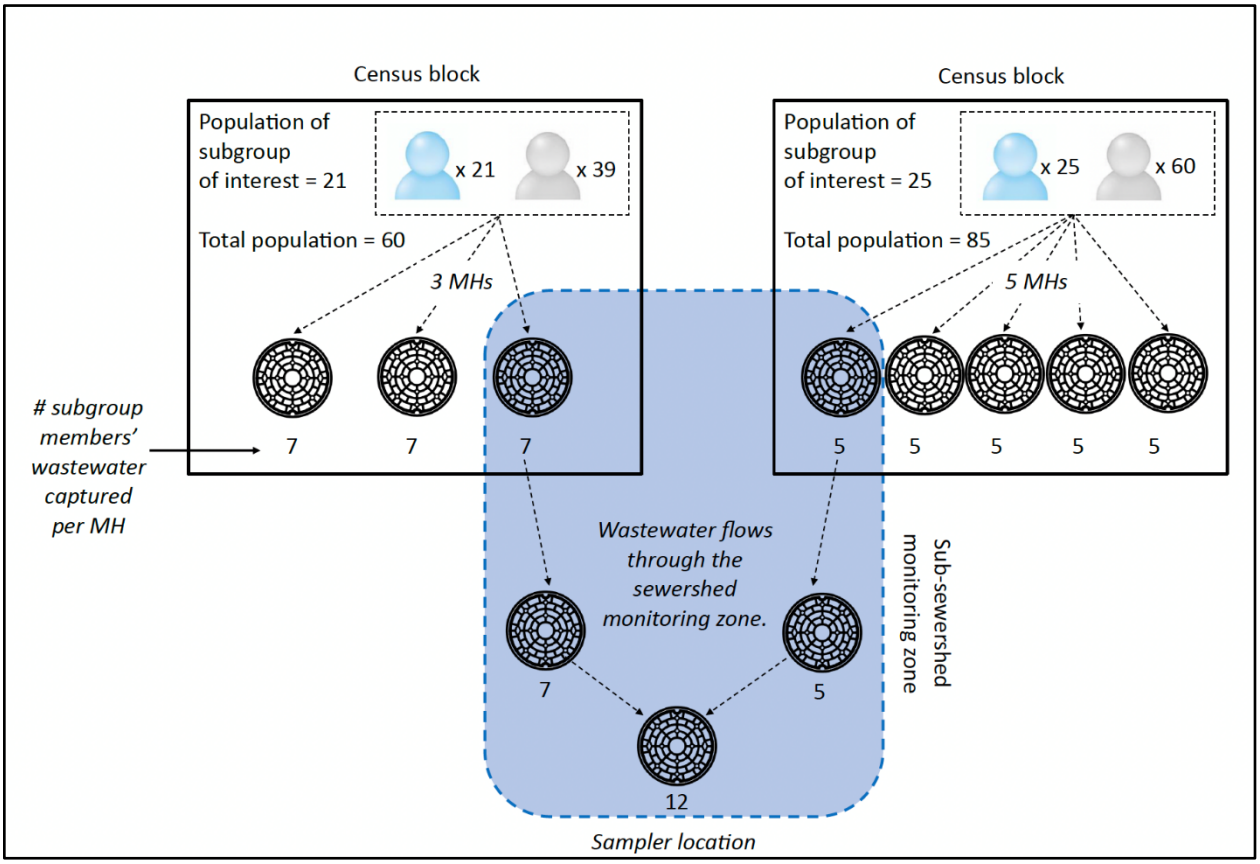


Figure 3. Simplified illustration of the probabilistic assignment method adopted from Safford et al. (2022), demonstrating how demographic data is distributed from a census block to a sub-sewershed zone.

Figure 3 illustrates how the locations of the MHs in the COD sewershed can be used to probabilistically assign demographic data from census blocks to sub-sewershed monitoring zones, whereas Safford et al. (2022) probabilistically assigned clinical case count data to sampling zones. This approach allows us to assess whether sampling locations were chosen in a way that appropriately represents subgroups within the overall population. The probabilistic assignment determines the expected subgroup population members whose wastewater is captured by sampling at a specific location. In this example, the sampler location depicted at the bottom covers a sub-sewershed monitoring zone that spans two census blocks. The census

block on the left has a total population of 60, of which 21 individuals belong to the subgroup of interest. There are three MHs in this census block. The census block on the right has a total population of 85, of which 25 individuals belong to the subgroup of interest. There are five MHs in this census block. Under the assumptions noted above, we calculated a predicted number of subgroup members captured by each MH in the census block by dividing the subgroup population by the number of MHs. In the case of the census block on the left, we divided the 21 individuals' wastewater contributions across all three MHs, resulting in a probabilistically assigned value of seven individuals. For the census block on the right, we split the 25 individuals' wastewater contributions across all five MHs, resulting in a probabilistically assigned value of five individuals. The wastewater flow was tracked through the connections between all the MHs in the city and summed at the sampler location. In this example, we obtained a predicted population of 12 people who belong to the subgroup of interest. Note that this is a simplified representation of the methodology behind the probabilistic tool. It is possible to obtain decimal resultant values using this method, but tool outputs can be rounded to whole numbers and still provide meaningful insight.

The output of the probabilistic tool for a single run can be generated in the format of a summary table, shown in Table 6. The table displays the number of community members from the subgroup of interest who are represented by the sample taken at a node under the given sampling zone boundaries. Additional columns denote sampling nodes by their MH identification (ID) number, and additional rows of the output table represent data from additional census blocks within the city. The values within each column under a MH ID number are the probabilistically assigned subgroup populations represented by the wastewater sample.

We ran the probabilistic distribution tool for each of the seven racial groups and each of the ten age categories. Then, the output table was filtered for the census blocks within each sub-sewershed zone to be able to calculate summary statistics by sub-sewershed zone rather than discrete census blocks.

Table 6. Example outputs from the probabilistic assignment tool for each sewershed node (truncated for visualization purposes).

Census block	Total subgroup population	Sewershed node							
		M16-011	N13-045	N11-062	N11-072	N12-066	O13-002	O20-001	…
061130104012002	13	0	0	0	0	0	0	13.0	…
061130107041016	8	0	4.0	0	0	0	4.0	0	…
…	…	…	…	…	…	…	…	…	…

We then compared the results of the probabilistic assignment tool to "manually derived" subgroup populations for each sub-sewershed zone. We obtained the manually derived population values by visually assigning the Census-reported subgroup population value to a given census block under the delineated zone boundaries. We calculated the absolute percent difference (APD) between the probabilistically assigned subgroup population value and the manually derived subgroup population value using the following equation:

$$APD = \left|\frac{P_A - P_M}{P_M}\right| * 100\%, \qquad \text{(Equation 2)}$$

where P_A is the probabilistically assigned population value of a subgroup within a sub-sewershed zone, and P_M is the manually derived population value of a subgroup within a sub-sewershed zone.

SUB-SEWERSHED ZONE AND COD WWTP ALIGNMENT

The following strategy was applied to aggregate sub-city level wastewater data to the city level. First, a set of population-weighted moving average values (PWMA) was calculated for

all fifteen sub-sewershed zones. This was done by multiplying the 10-day trimmed moving average N/PMMoV values by the sub-sewershed zone population and dividing those values by the total city population. These PWMA values were then plotted, as can be seen in Figure 4. A city-aggregated N/PMMoV metric was determined by summing the PWMA values for the sub-sewershed zones and creating a set of combined PWMA values. These values were then plotted against the unweighted COD WWTP moving average values (See Figure 5). Note that, in this summation, SR-C3 was not included since it is encompassed within the SR-C2 sub-sewershed zone. To assess the strength of the correlation between the combined sub-sewershed values and the city-level wastewater data, we calculated and reported the Spearman's Rank correlation coefficient. This metric was chosen because it measures strength and direction of any monotonic relationship rather than solely a linear one (*Spearman Rank - an Overview | ScienceDirect Topics*, n.d.).

To quantify the error associated with the cumulative population-weighted moving average values, we calculated the mean absolute error (MAE) for each sub-sewershed zone's set of population-weighted moving averages, using the following equation:

$$MAE = \frac{1}{n} * \sum_{i=1}^{n} |R_i - PWMA_i|, \qquad \text{(Equation 3)}$$

where R_i is the raw wastewater data value for a given sub-sewershed zone, $PWMA_i$ is the associated population-weighted moving average value for the zone, and n is the total number of data points. Since these resultant error values are then added together to generate the cumulative population-weighted moving average values, error propagation must be considered (*Propagation of Error*, 2013). The additive formula for error propagation was used to combine these uncertainties:

$$Error = \sqrt{(MAE_1)^2 + (MAE_2)^2 + (MAE_3)^2 + \cdots + (MAE_n)^2}, \quad \text{(Equation 4)}$$

where MAE_i is a given mean absolute error value in the set of error values for a sub-sewershed zone.

RESULTS AND DISCUSSION

PROBABILISTIC ASSIGNMENT BY RACE AND AGE

One aspect of considering health equity in the design of sub-city sampling strategies for wastewater monitoring includes evaluating the percent of diverse subgroup populations represented by a set of sub-sewershed zones selected. Yet demographic data needed to perform this assessment are available for census blocks, which do not directly align with sub-sewershed zones. We compare two approaches—one manual and one probabilistic—to determine subgroup populations in sub-sewershed zones using census block data. The APD values serve as a metric for detecting differences between the subgroup population values generated using manual derivation and probabilistic assignment within each sub-sewershed. An APD close to 0% indicates that there is minimal difference between values obtained manually or probabilistically when devising a monitoring strategy, while a value closer to 100% indicates a larger difference between the two methods. The APD values can be assessed spatially in terms of geography (sub-sewershed zones) and demographics (race and age).

We display the absolute percent difference for the two approaches for each sub-sewershed zone in the City of Davis as choropleth maps for each subgroup category of race (Figure 6) and age (Figure 7). Dark red zones signify high percent difference between the two approaches, while lighter yellow zones indicate low percent differences. The APDs obtained

from the probabilistic tool output are also listed in Table 6 and Table 7. High APDs suggest that there are large boundary effects for the subgroup population of interest. In these cases, population distributions overlapping multiple census blocks should be considered when evaluating the overall coverage of subgroup populations in a city's sampling strategy.

APDs values across all sub-sewershed zones in Davis for race categories (Table 9) ranged from approximately 0% to 43%. Choropleth maps of APD by race category in each sub-sewershed zone (Figure 6) display relatively low percentage differences in SR-A, SR-D, and SR-I. SR-B and SR-E showed consistently higher APD values. We used 30% as a threshold to flag notable differences between probabilistically assigned subgroup populations and manually derived population values in each sub-sewershed zone. Overall, greater percent differences were observed more often for minority subgroup populations compared to White populations. This highlights how boundary effects can be significant for subgroups present in low absolute numbers. For the White population, only SR-E had a value above the 30% threshold (34.9%). For the Black or African American population, APDs above 30% were seen in SR-B (36.4%) and SR-E (33.8%). For the American Indian and Alaska Native population, APDs surpassed the threshold in SR-E (35.7%), SR-F (40.8%), and SR-H (42.9%). For the Asian population, APDs were above the threshold in SR-B (34.6%). For the Native Hawaiian and Other Pacific Islander populations, APDs above 30% were observed in SR-E (42.9%) and SR-G (33.3%). For populations of individuals belonging to another race category, values above the APD threshold were seen in SR-B (38.7%), and for multiracial populations, these values were seen in SR-E (30.3%).

APDs across all sub-sewershed zones in Davis for age categories (Table 10) ranged from approximately 0% to 60%. Choropleth maps of APD values by age category in each sub-

sewershed zone (Figure 7) display the relatively low percent difference values in SR-A, SR-D, SR-H, and SR-I. Consistently higher APD values were observed in SR-B, SR-C, and SR-E. However, we do see lower APD values in some of these zones for age categories with older populations, which is significant since these are populations who are more susceptible to severe health outcomes (*COVID-19 Death Data and Resources - National Vital Statistics System*, 2023). For the population under age 5 and those ages 5-17, only SR-E had an APD above the 30% threshold (34.0% and 36.3%, respectively). For the population ages 18-34, APDs above 30% were seen in SR-B (36.3%). For those in the 35-49, 50-59, 60-64, 70-74, 75-79, and 80 and above age categories, such values were seen only in SR-E (32.6%, 32.8%, 42.9%, 47.1%, 60.0%, and 43.9%, respectively). For those ages 65-69, APDs above 30% were seen in SR-C (31.1%) and SR-E (45.0%).

More populous zones like SR-A demonstrated consistently lower APDs, while zones like SR-E, which has a relatively smaller population, had the highest APDs across all sub-sewershed zones when analyzed by race and age. The only zones where no difference was observed between the probabilistic assignment and manual derivation methods (APD=0%) were SR-D and SR-H, two zones with relatively lower populations compared to the other seven zones studied in this analysis. In subsequent analysis to provide examples of adaptive sampling frameworks, we selected to use the probabilistic assignment tool to assess the representation of subgroup populations across Davis under different planning scenarios.

SUB-SEWERSHED ZONE AND COD WWTP ALIGNMENT

In the selection of sewershed nodes for wastewater-based disease surveillance, a decision-maker may also want to consider how disease dynamics in sub-sewershed zones

compare to overall city disease trends. Sampling frameworks could then be adapted, for example, to capture sub-sewershed zones with highly differing disease dynamics than the city overall. Alternatively, sampling frameworks that adequately capture overall trends for the city may be prioritized. Here, we evaluate whether wastewater surveillance data collected at the sub-sewershed level in Davis correlated in aggregate with the COD WWTP wastewater surveillance data. We then assess how the correlations would be impacted when downscaling to fewer nodes in an example scenario planning exercise for adaptive sampling. Analysis of this type can also be useful to identify anomalies at the sub-city scale relative to the city-level data, revealing potentially important health vulnerabilities in different city regions.

Individual sub-sewershed zones exhibited variability in the magnitude and timing of wastewater signals (Figure 4). As expected, the propagated error for the aggregated sub-sewershed zone data ($6.48 * 10^{-3}$) was much higher than that of the COD WWTP moving averages ($3.60 * 10^{-4}$) due to the greater number of operations performed to generate the data. Nevertheless, the cumulative population-weighted mean average of the sub-sewershed zones exhibited similar magnitudes and patterns compared to the COD WWTP moving average wastewater data (Figure 5). The Spearman's Rank correlation coefficient over the study period was 0.909, with a statistically significant positive correlation (p-value of $5.88 * 10^{-28}$). The strongest correlations were observed between February 2022 and May 2022. Decoupling over the study period may be attributed in part to the fact that during the summer and early winter months, the population in Davis fluctuates due to the movement of students in and out of the city. A Spearman's Rank correlation coefficient was also calculated between each sub-sewershed zone and the COD WWTP (listed in Table 7). Correlation coefficients ranged from

approximately 0.732 to 0.935, and all p-values were less than 0.05. This suggests that many combinations of sub-sewershed zones may offer reasonable representation of wastewater disease dynamics at the city-level, though site-specific fluctuations from sub-sewershed zone data can still be used to target public health responses. Moreover, wastewater data at the sub-sewershed level can be used to alert health professionals to potential surges in certain regions within a city. For instance, wastewater data from four sub-sewershed zones (SR-A, SR-B3, SR-C1, and SR-C2) tended to rise earliest amongst all the zones, pointing to potential regions for early health interventions. In fact, HDT actively utilized wastewater data from sub-sewershed zones at the time of the data collection as a component to a multi-faceted strategy for precision public health (H. Safford et al., 2022).

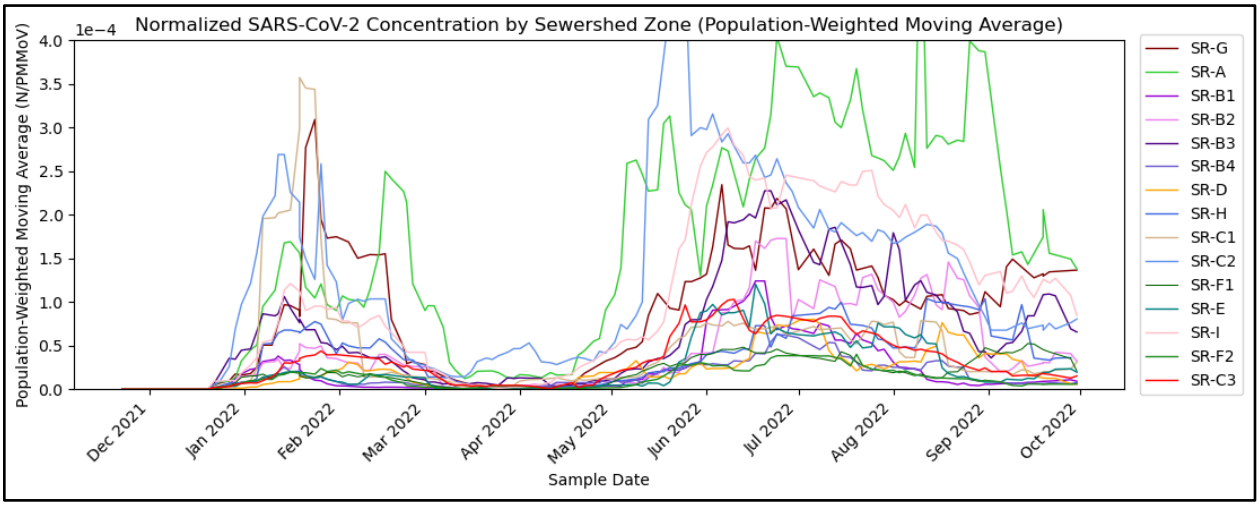


Figure 4. Disaggregated population-weighted 10-day right-aligned trimmed moving average values. Each line represents a sub-sewershed zone monitoring location.

Table 7. Correlation of wastewater data for each sub-sewershed zone (population-weighted moving average) with the overall COD WWTP data (moving average).

Zone Name	Correlation Coefficient	P-Value
SR-A	0.770	4.19e-15
SR-B1	0.935	8.81e-33
SR-B2	0.836	1.24e-19
SR-B3	0.927	4.69e-31
SR-B4	0.922	3.26e-30
SR-C1	0.793	1.69e-16
SR-C2	0.890	3.22e-25
SR-D	0.734	3.15e-13
SR-E	0.877	1.13e-23
SR-F1	0.877	1.24e-23
SR-F2	0.870	6.78e-23
SR-G	0.738	2.20e-13
SR-H	0.733	3.88e-13
SR-I	0.866	1.92e-22

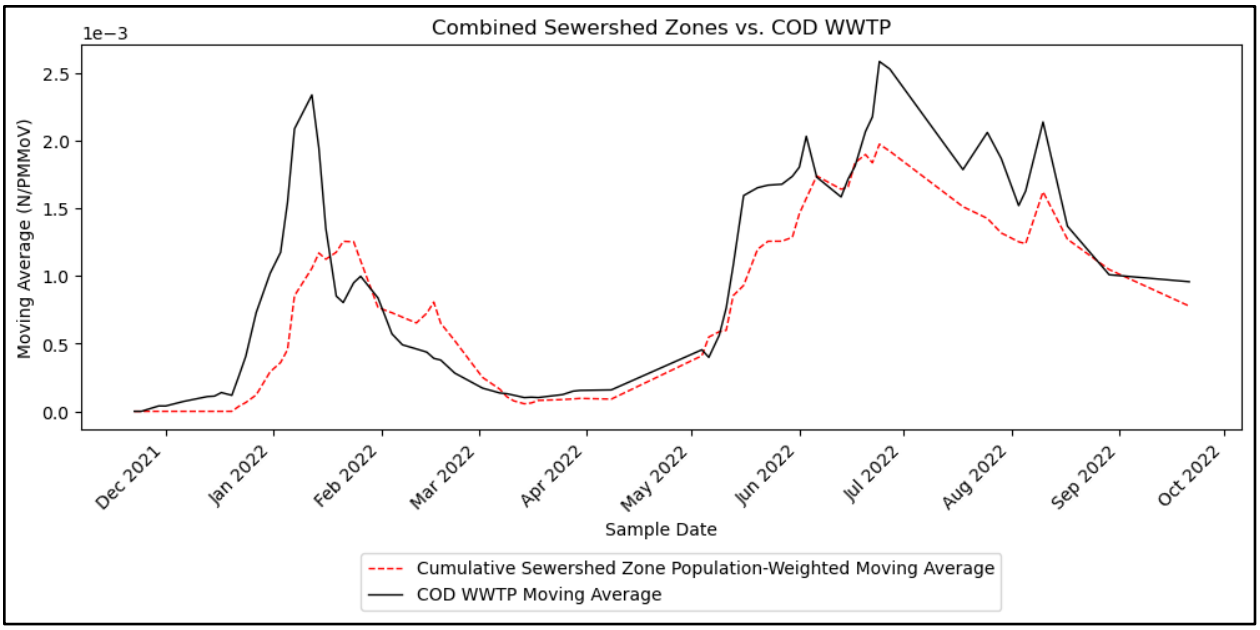


Figure 5. Aggregated wastewater data from the City of Davis sub-sewershed zones using a right-aligned population-weighted moving average (red, dashed line) and wastewater data from the City of Davis WWTP using a right-aligned moving average (black, solid line).

SCENARIO PLANNING FOR ADAPTIVE SAMPLING

Here we offer an example of how subgroup population assignment and wastewater data aggregation approaches can be used in scenario planning to design sub-sewershed sampling strategies for a city operating under resource constraints (e.g., logistical or economic). We considered four examples of resource-constrained conditions that necessitate scale-back of the number of sampling sites compared to baseline. Baseline was defined as the sampling framework established for this study, which covers a population of 63,826 (95.5% of the Census-reported total population in 2020). Scenarios 1 and 2 evaluate impacts for a minor (~25%) reduction in the number of sampling nodes, while Scenarios 3 and 4 evaluate a reduction of sampling nodes by approximately 50%. In each scenario we evaluated the impact of sampling node selections on (1) the representation of the >65-year-old population (who comprise about 14.5% of the COD population), (2) the representation of the Black or African

American populations (who comprise about 2.31% of the COD population), and (3) correlations between aggregated sub-sewershed zone wastewater data with city-level data.

Tradeoffs may be observed when covering different demographics, and downscaling sampling may impact the representativeness of the data to the city overall. In this example, we stratified the sub-sewershed populations by age, focusing on the high-risk population of individuals ages 65 and older. We observed how modifying the sampling regime by dropping either approximately 25% or 50% of sampling sites affects population representation for subgroups disproportionately impacted by COVID-19 (elderly populations and Black or African American populations). In Scenarios 1 and 3, the sampling sites removed were chosen at random. In Scenarios 2 and 4, the sampling sites removed were selected such that the coverage of the >65-year-old population was prioritized. In other words, the nodes that were removed corresponded to sub-sewershed zones with the smallest subgroup populations in the city. In Table 8, we report four values for each scenario: the percentage of the total city population covered by the scenario (i.e., not including the populations in the zones whose nodes were removed), the percentage of the >65-year-old subgroup population covered by the sampling regime, the percentage of Black or African American subgroup population covered under that same regime, and a Spearman's Rank correlation coefficient that reports the strength of correlation between the cumulative PWMA values and COD WWTP values without the inclusion of the removed zones.

For the baseline scenario, all the sub-sewershed nodes monitored in this study were included. The probabilistically assigned subgroup population values were used to determine the total subgroup populations represented within the proposed sampling regime. The baseline

scenario of sub-sewershed sites was designed to capture wastewater from as much of the sewershed as possible while providing resolution within the city boundaries to inform public health interventions. This objective was achieved, as demonstrated by 95.5% coverage of the population overall, 86.9% coverage of the >65-year-old population, and 82.9% coverage of the Black or African American populations within the city.

In Scenario 1, we examined the city-level effect of artificially pausing sampling at 25% of the sampling sites, chosen at random (SR-A, SR-D, SR-F1, and SR-F2 were removed). In this scenario, the number of sampling nodes was reduced from 15 to 11. When doing so, we observed that the wastewater contributions of approximately 57.5% of the >65-year-old population would be captured by the sampling regime (total subgroup population in the included zones divided by the total subgroup population in Davis). The wastewater contributions of about 70.3% of the Black or African American population would be captured. The aggregated wastewater signal was minimally affected by removing the excluded zones' sampling nodes. The Spearman's Rank correlation coefficient of the cumulative PWMA values (excluding SR-A, SR-D, SR-F1, and SR-F2) to the COD WWTP moving average values remained high (0.907 with a p-value of $1.19 * 10^{-27}$).

In Scenario 2, we examined the city-level effect of pausing sampling at 25% of the sampling sites, chosen to maximize coverage of >65-year-old populations (SR-D, SR-E, SR-F1, and SR-F2 were removed). Like Scenario 1, the number of sampling nodes was reduced from 15 sites to 11. As expected, coverage of the >65-year-old population improved relative to Scenario 1, such that 80.5% of the >65-year-old population would be captured in this sampling regime. Unexpectedly, the coverage of Black or African American populations also improved in this

scenario, for which about 89.7% of the population would be captured. The aggregated wastewater signal, excluding SR-D, SR-E, SR-F1, and SR-F2, compared to the COD WWTP moving averages resulted in a Spearman's Rank correlation coefficient of 0.900 with a p-value of $1.19 * 10^{-26}$, reflecting a minimal change from the baseline case.

In Scenario 3, we examined the effect of pausing sampling at approximately 50% of the sampling sites, chosen at random (SR-B1, SR-B2, SR-B3, SR-B4, SR-H, and SR-I were removed). In this scenario, the number of sampling nodes was reduced from 15 to 9. When doing so, we observed that the wastewater contributions of approximately 51.1% of the >65-year-old population would be captured by the sampling regime. The wastewater contributions of about 67.5% of the Black or African American population would be captured. The Spearman's Rank correlation coefficient of the cumulative PWMA values (excluding SR-B1, SR-B2, SR-B3, SR-B4, SR-H, and SR-I) to the COD WWTP moving average, while lower compared to the baseline, remained strong (0.875 with a p-value of $1.87 * 10^{-23}$).

In Scenario 4, we examined the effect of pausing sampling at about 50% of the sampling sites, chosen to maximize coverage of >65-year-old populations (SR-D, SR-E, SR-F1, SR-F2, SR-G, and SR-H were dropped). Like Scenario 3, the number of sampling nodes was reduced from 15 sites to 9. Coverage of both >65-year-olds and Black or African American individuals improved relative to Scenario 3. In the sampling regime for Scenario 4, wastewater contributions of about 67.2% of the >65-year-old population and 76.7% of the Black or African American population would be captured. The aggregated wastewater signal also correlated more strongly with the city overall. Excluding SR-D, SR-E, SR-F1, SR-F2, SR-G, and SR-H in the cumulative PWMAs and comparing them to the COD WWTP moving averages resulted in a Spearman's Rank correlation

coefficient of 0.917 with a p-value of $2.83 * 10^{-29}$. This suggests that the removed zones deviate from the city-level results, demonstrating the utility of wastewater monitoring at the sub-sewershed scale to illuminate site-specific variations in wastewater virus concentration.

Table 8. Scenario planning example to adapt sampling strategies under resource constraints and differing priorities.

Scenario	Description	% of Total City Population Represented	% of >65-Year-Old Subgroup Population Covered	% of Black or African American Subgroup Population Covered	Wastewater Signal Correlation
Baseline	Includes all sewershed nodes monitored in this study	95.5%	86.9%	82.9%	0.909
1	Baseline minus 25% of sites (randomly selected)	69.1%	57.5%	70.3%	0.907
2	Baseline minus 25% of sites (prioritized to maximize coverage of >65-year-old population)	84.9%	80.5%	89.7%	0.900
3	Baseline minus 50% of sites (randomly selected)	61.8%	51.1%	67.5%	0.875
4	Baseline minus 50% of sites (prioritized to maximize coverage of >65-year-old population)	70.5%	67.2%	76.7%	0.917

The scenario planning example outlined above revealed that removing sampling sites had little effect on the wastewater signal correlation across all the scenarios examined. As expected, demographic comparison revealed that greater >65-year-old coverage is achieved when scale-back of sampling is methodical rather than random. Prioritizing coverage of subgroup populations of interest consistently yields higher levels of coverage, as seen in Scenarios 2 and 4. However, the example demonstrates that randomly selecting sample sites to drop may be appropriate situationally, especially if the scale-back efforts are not drastic (e.g., dropping less than 25% of sites). We note that prioritizing certain subgroups will result in varied representation of other subgroups, as seen in this example. We observe how the Black or

African American subgroup population (2.31% of the total COD population) representation decreased as a result of the scale-back, to be expected. It is ultimately up to decision-makers to determine whether the resultant percentages are sufficiently representative of the city population, else run the risk of overlooking a portion of the population. However, when making decisions related to improved health interventions within a city, it is simultaneously important to avoid targeting specific populations and creating additional stigmas associated with disease transmission (Holm, Osborne Jelks, et al., 2023). There may be cases in which regional specificities contribute more heavily to decision-making. In such cases, there is a need to balance the perspectives of officials at the city and county level, as well as medical staff from public and private organizations. Stakeholders may factor in the demographic representation fluctuations seen in specific sub-sewershed zones into decision-making instead of immediately opting for the highest subgroup population percentage seen at the city level within a specific modified sampling regime. The example provided in this study showcases the utility of sub-sewershed demographics analysis to gain more detailed insight into how representation varies depending on the scenario. While the city-level wastewater data may not suggest any data gaps, looking specifically at the sub-sewershed demographics reveals that there is some difference between choosing one scenario versus the other. Therefore, this example demonstrates how investigating sub-sewershed trends can foster more informed decision-making in city-wide monitoring efforts, whether that involve selecting sampling locations in a way that is representative of the population, inspecting how the selected zone populations' wastewater data aggregates to city-level data, or gaining a better understanding of population trends within a specific subset (e.g., higher age brackets and/or non-White populations). Sub-

sewershed analyses demonstrate the possible ways that sampling site selection can be adaptive to the unique location where monitoring efforts are being undertaken, while avoiding diverting monitoring away from high-risk populations.

LIMITATIONS

We acknowledge several limitations to this study and approach. First, the approach relies on the availability of and integrity of the underlying Census data. As mentioned previously, few demographic factors had sufficient data available for this analysis, and increased margins of error were reported for the 2020 Census data (Bureau, n.d.-b). The Census Bureau has also incorporated statistical noise into its datasets in its efforts to implement differential privacy (*Disclosure Avoidance and the 2020 Census: How the TopDown Algorithm Works*, n.d.). Therefore, the population values used in this study may not be representative. However, noise infusion is beneficial for addressing privacy concerns, as it maintains the essence of the data while ensuring that the data cannot be traced to an individual. When working with Census data, especially at its highest granularity, privacy protection is crucial so that public trust is not undermined through the mishandling of private information, and health disparities are not subsequently reinforced.

Second, it is important to note that the approach may need to be tailored when extended to other settings, as differences will arise from regional specificities. This study specifically examines the monitoring protocol as it was conducted in Davis, a suburban town with institutional support for its rigorous wastewater testing program during the pandemic. Though the tool presented in this study should be applicable to other settings, differences

including, but not limited to, sewer system layout, resource constraints, and population density, may require additional analysis.

Third, this study does not account for the mobility of populations, as wastewater surveillance efforts are place-based. Reported metrics may not be taking into account individuals' place of residency or work (Kadonsky et al., 2023). Additionally, Davis comprises a large student population, so some sub-sewershed zones are likely to include more students who typically move frequently. Therefore, Census-related fluctuations and variability within the data may be higher in these areas compared to regions that are primarily inhabited by long-term residents. Extending the approach outlined in this study to evaluate data in other regions and cities would offer opportunities for cross-comparisons.

Fourth and finally, this study analyzes race and age separately, as though their distribution of impact is mutually exclusive. In reality, there are populations (e.g., Black or African American population above the age of 65) that are especially vulnerable to COVID-19. Evaluating health equity on the basis of single characteristics is inherently limiting, as it may be devoid of other important context and thereby risks homogenizing marginalized groups.

POSSIBLE EXTENSIONS

Several reasonable extensions of this project arise as we continue to understand community transmission of SARS-CoV-2 and other pathogenic viruses and explore ways that we can inform equitable public health decision-making. While this study focuses on community transmission at the city-level, population representation should also be assessed at county and regional levels (Medina et al., 2022). Race and age were the two demographic characteristics analyzed in this study, though there are many other attributes that can provide insight into the

health equity implications of a wastewater surveillance program, such as socioeconomic status, healthcare coverage, immigrant status, native language, and educational attainment. Future analyses could consider these factors at appropriate scales for available data—the American Community Survey, for instance, offers a breadth of demographic data at broader scales (e.g., census block group or tract) (Bureau, n.d.-a). Moreover, since health disparities are driven by a combination of all these factors, future analyses can take intersectionality into consideration by analyzing the cumulative impact of several characteristics on health outcomes (e.g., using quantitative analytical methods that adjust for covariates) and developing social vulnerability indices that can inform sampling location selection processes (Guan et al., 2021). For a well-balanced distribution of sampling nodes, optimization models can be developed to consider the users' parameters of interest. Researchers are developing objective functions that account for myriad factors, including the population served, the population density, spatial coverage, social vulnerability, and dissimilarity of wastewater signals, in the design of sampling schemes at broader geographic scales (Nuño, 2024). These approaches can be applicable at varying scales of deployment for wastewater-based disease surveillance (e.g., building, neighborhood, city, county, or regional).

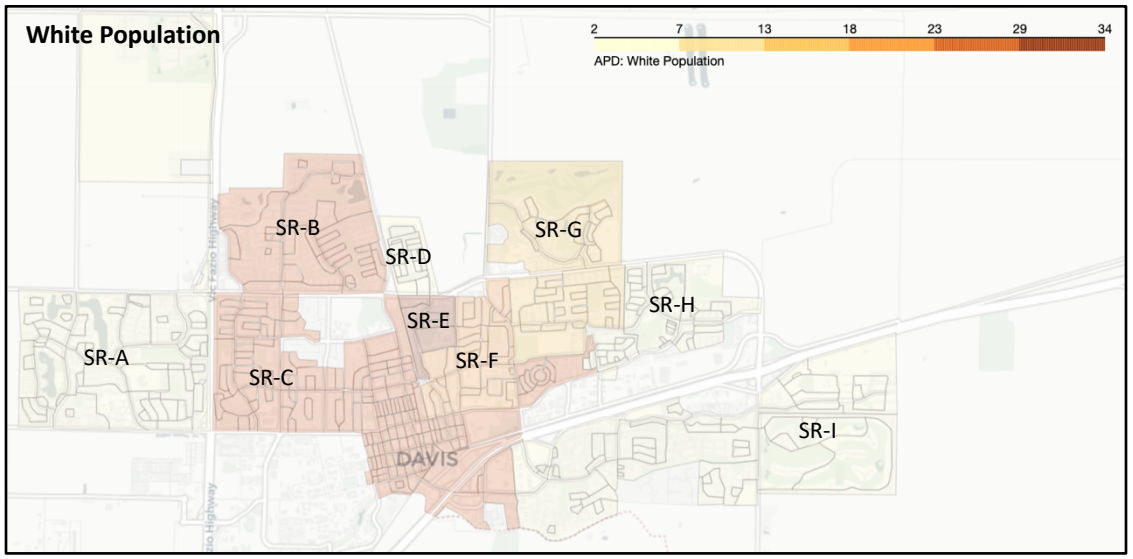
CONCLUSION

A successful wastewater-based disease surveillance program should aim for equitable representation of vulnerable groups within the sampling design, while preserving anonymity of the sampled populations. When sampling frameworks are designed appropriately and equitably, wastewater data can inform health interventions that address disparities faced by underserved communities. The probabilistic assignment approach demonstrated in this study

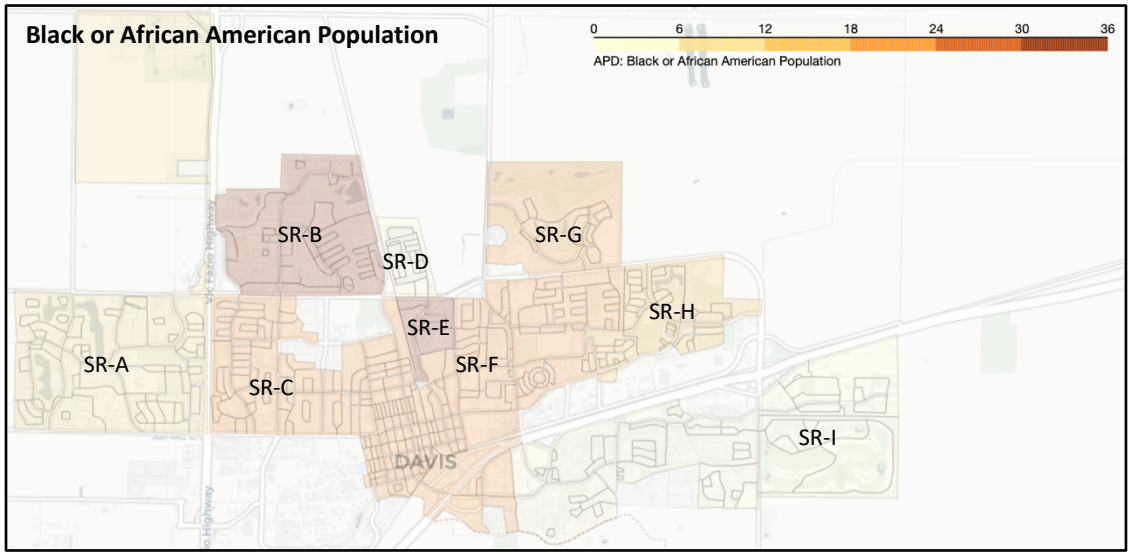
offers a way to determine the distribution of impact of a planned (or existing) change in sampling regimes at a sub-sewershed level, thus facilitating equitable sampling design amidst shifting priorities and variable conditions. The approach is scalable and can be adapted to incorporate factors beyond those reported in this study.

This study also demonstrates the utility of using probabilistic distribution in resource-constrained settings for scenario planning. If only a certain number of autosamplers can be deployed in an area, the procedure outlined in this study could be used to determine favorable locations to place samplers to optimize a chosen measure (e.g., capture a larger proportion of the population over the age of 65, who are more vulnerable to serious illness from COVID-19). Then, sampling locations can be chosen strategically to ensure that majority of the population is being covered through sample collection at those sites. Ultimately, using this tool can help stakeholders adapt their sampling strategies to the local social landscape, thereby incorporating equity-based criteria into public health policies and striving towards an equitable sampling regime.

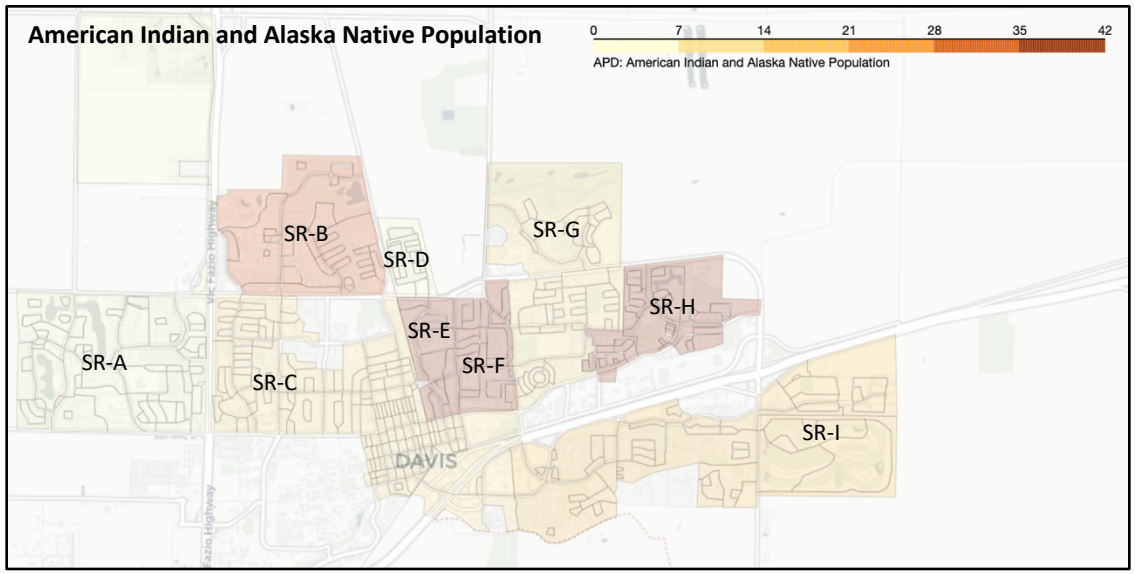
a)


White Population
2 7 13 18 23 29 34
APD: White Population
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

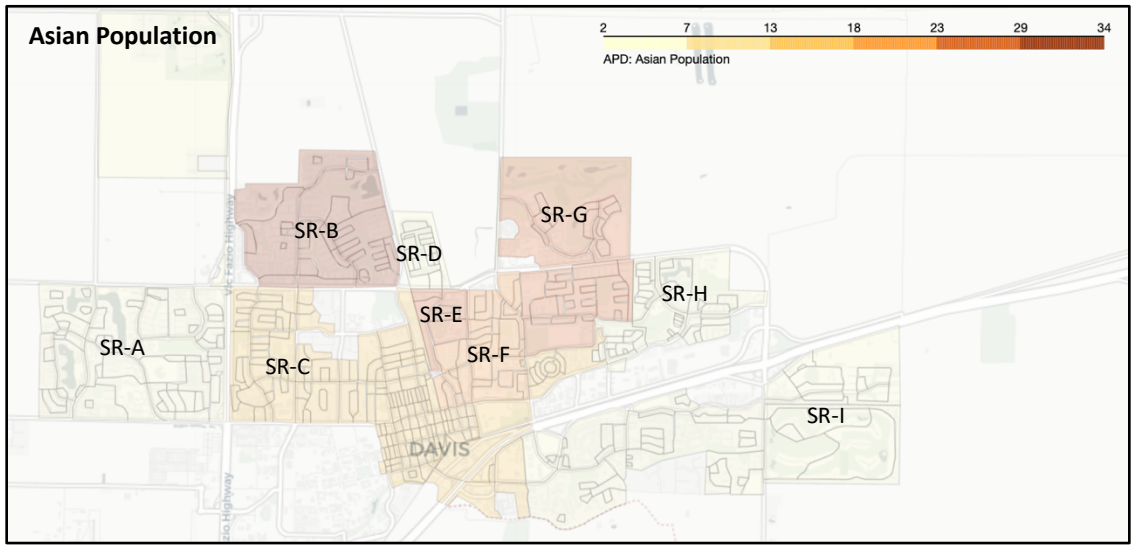
b)


Black or African American Population
0 6 12 18 24 30 36
APD: Black or African American Population
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

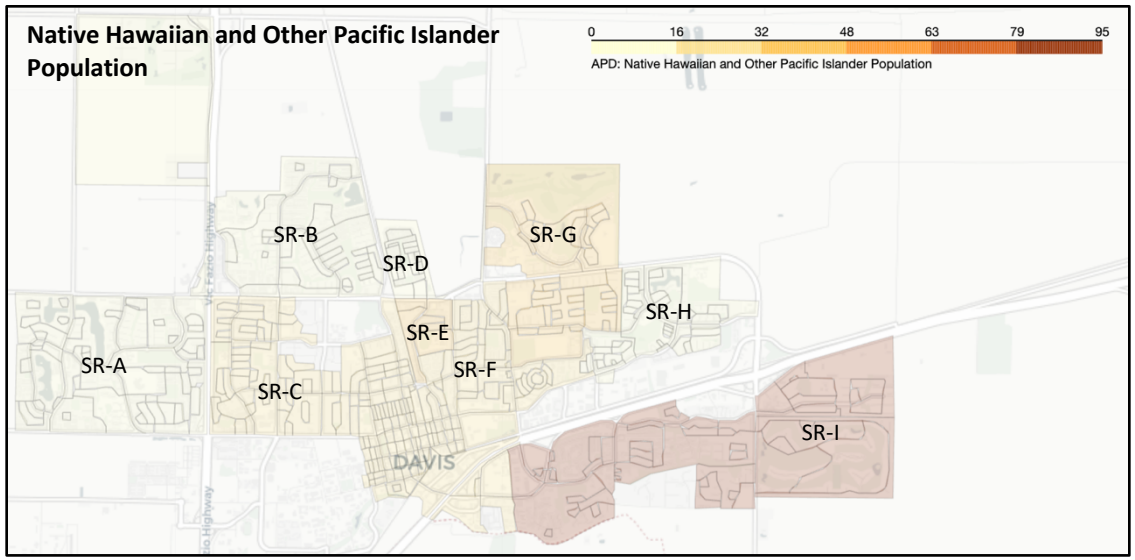
c)


American Indian and Alaska Native Population
0
7
14
21
28
35
42
APD: American Indian and Alaska Native Population
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-C
SR-F
SR-I
DAVIS

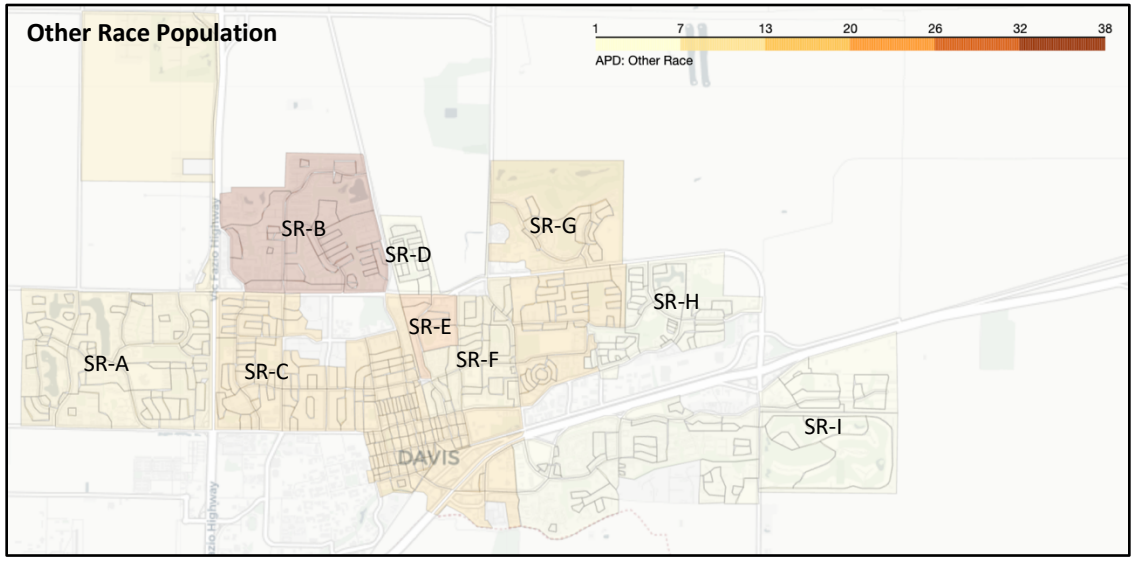
d)


Asian Population
2
7
13
18
23
29
34
APD: Asian Population
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-C
SR-F
SR-I
DAVIS

e)


Native Hawaiian and Other Pacific Islander Population
0
16
32
48
63
79
95
APD: Native Hawaiian and Other Pacific Islander Population
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

f)


Other Race Population
1
7
13
20
26
32
38
APD: Other Race
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

g)

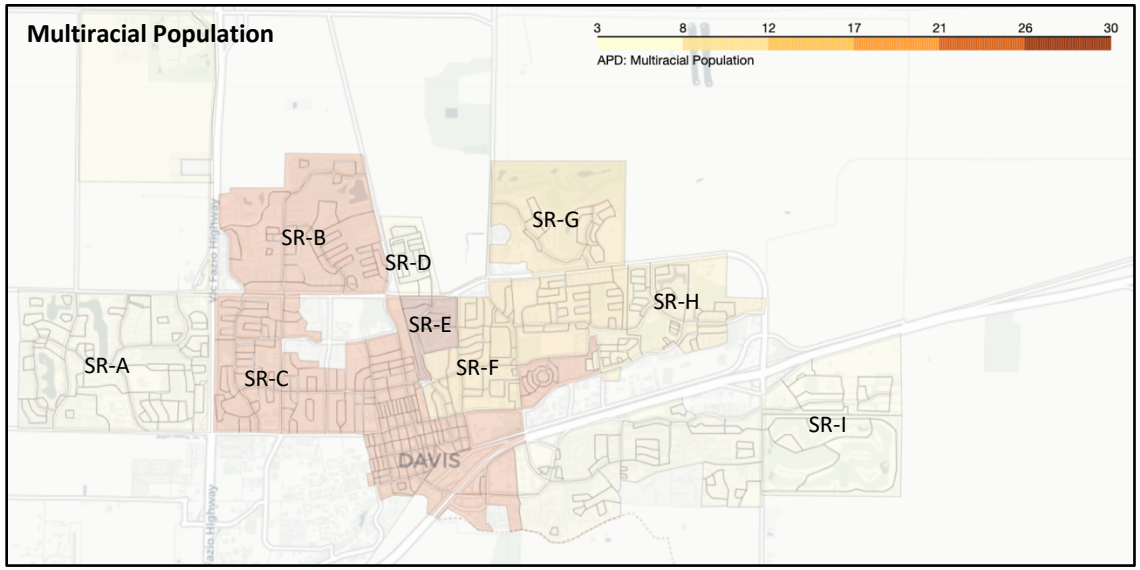
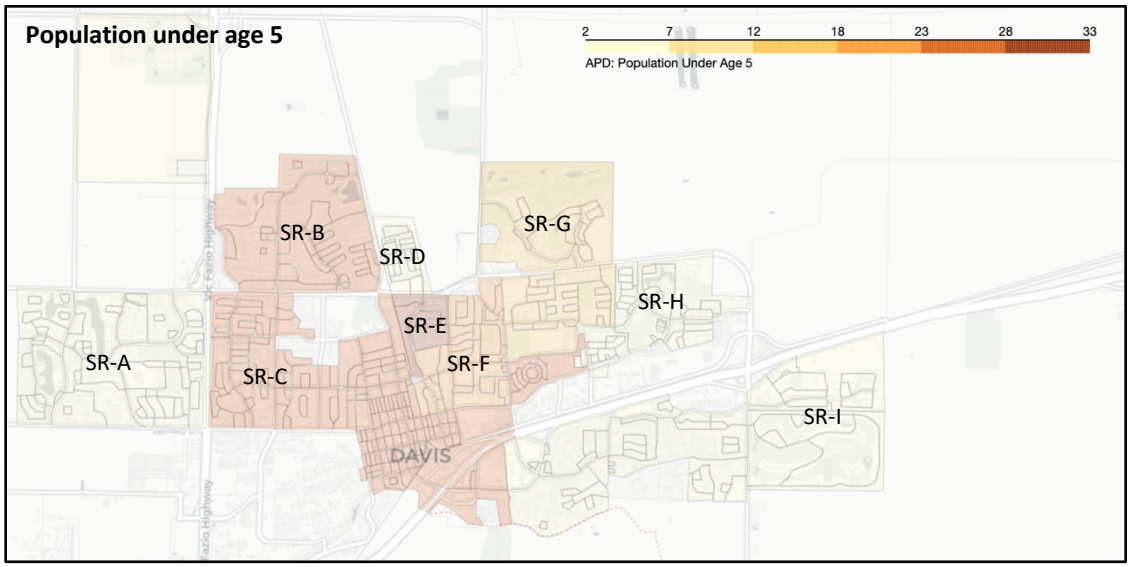


Figure 6. Choropleth maps of absolute percent differences (APDs) between probabilistic and manual assignment of census block data for race into sub-sewershed zones. Results are shown for each racial subgroup population captured through HDT wastewater sampling, by sub-sewershed zone. Stratifications included (a) White population, (b) Black or African American population, (c) American Indian and Alaska Native population, (d) Asian population, (e) Native Hawaiian and Other Pacific Islander population, (f) Other race population, and (g) Multiracial population.

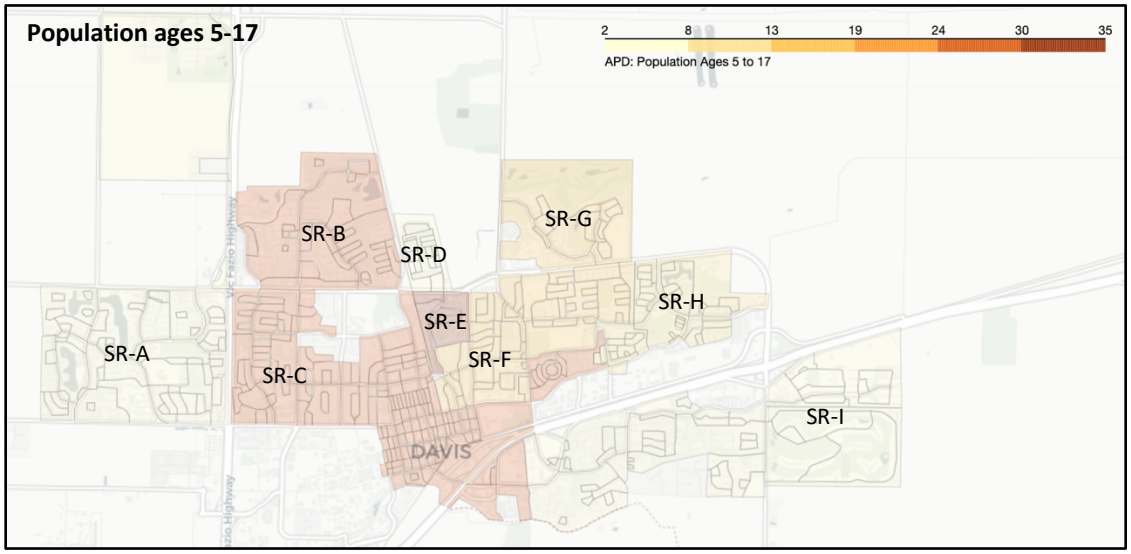
Table 9. Absolute percent differences (APDs) between probabilistic and manual assignment of census block data for race into sub-sewershed zones.

Zone Name	White	Black or African American	American Indian and Alaska Native	Asian	Native Hawaiian and Other Pacific Islander	Other	Multiracial
SR-A	3.8	6.4	3.6	5.7	5.0	10.1	6.2
SR-B	25.8	36.4	28.6	34.6	16.7	38.7	24.6
SR-C	27.4	20.3	12.8	15.4	26.9	18.9	25.1
SR-D	3.5	0.0	0.0	2.2	0.0	0.9	3.0
SR-E	34.9	33.8	35.7	27.0	42.9	22.6	30.3
SR-F	21.5	23.7	40.8	20.4	12.5	11.8	16.1
SR-G	15.5	20.0	11.1	25.3	33.3	19.4	15.8
SR-H	6.4	14.9	42.9	3.8	0.0	6.6	10.0
SR-I	2.8	5.6	17.1	3.6	4.0	4.7	5.3

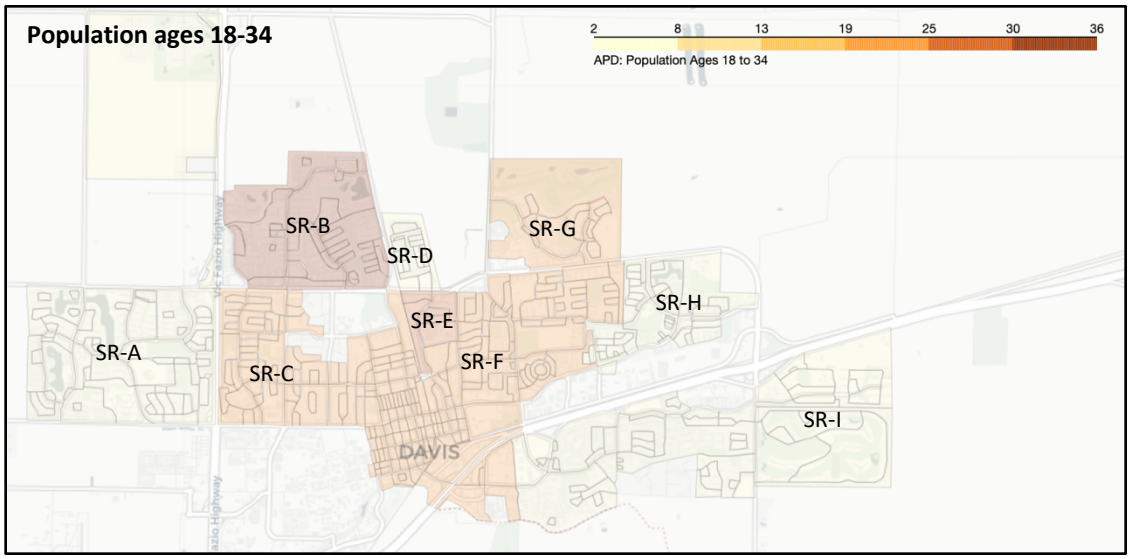
a)


Population under age 5
2 7 12 18 23 28 33
APD: Population Under Age 5
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-C
SR-F
SR-I
DAVIS

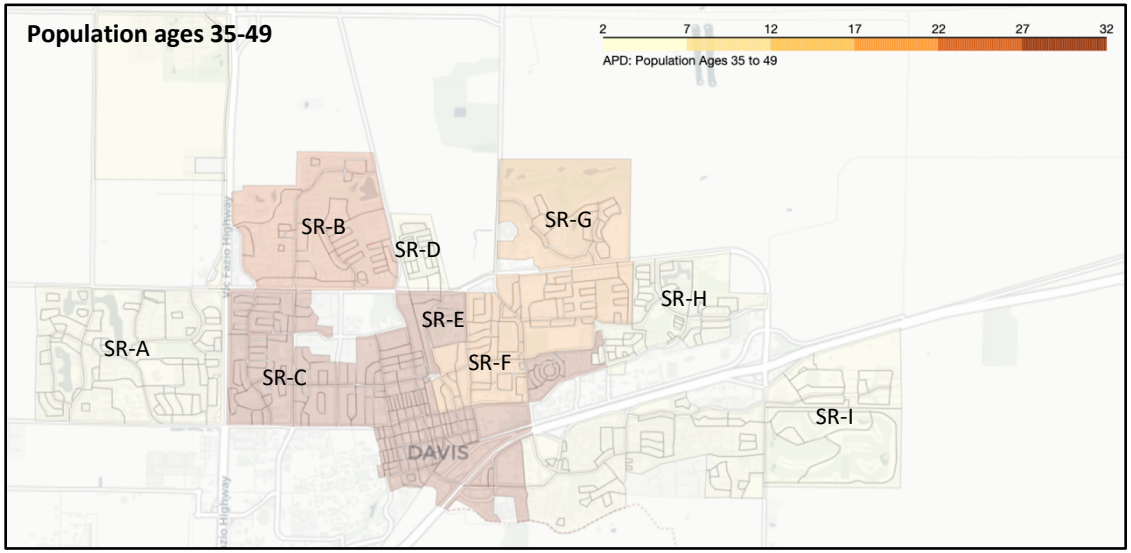
b)


Population ages 5-17
2 8 13 19 24 30 35
APD: Population Ages 5 to 17
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-C
SR-F
SR-I
DAVIS

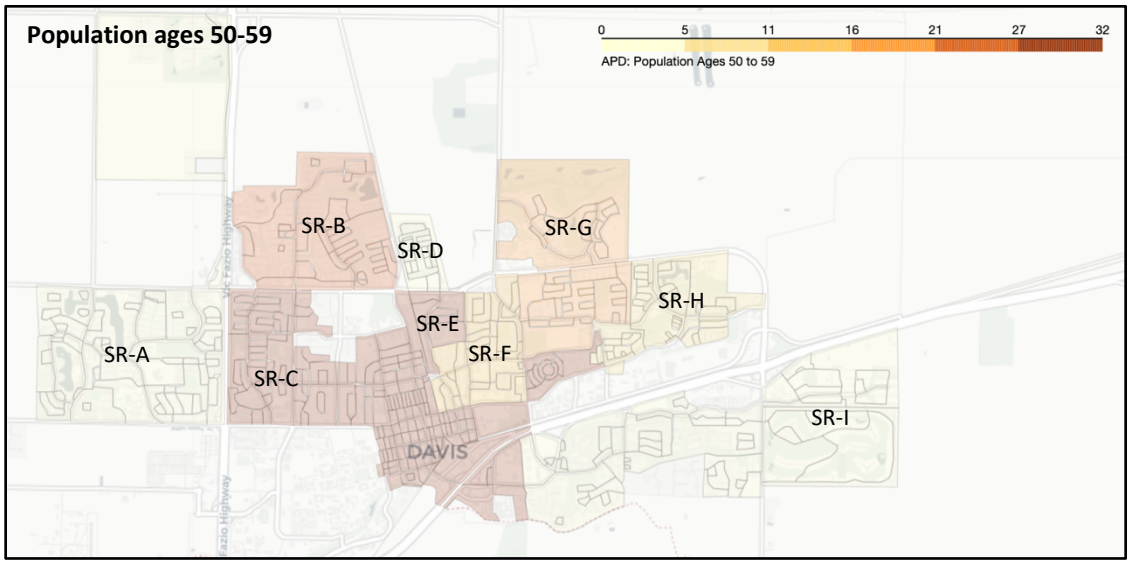
c)


Population ages 18-34
2 8 13 19 25 30 36
APD: Population Ages 18 to 34
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

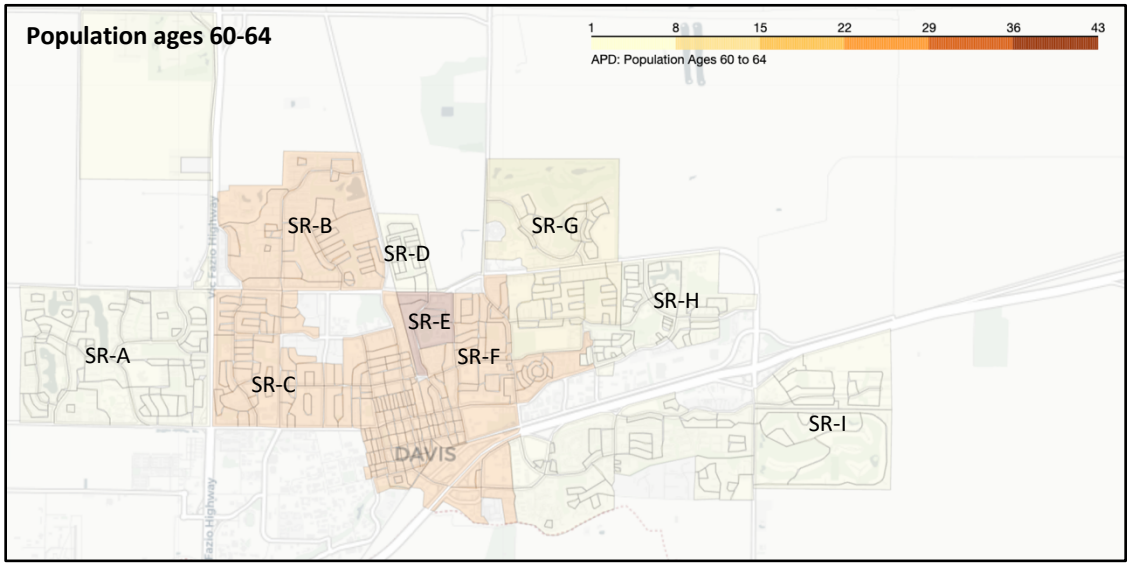
d)


Population ages 35-49
2 7 12 17 22 27 32
APD: Population Ages 35 to 49
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

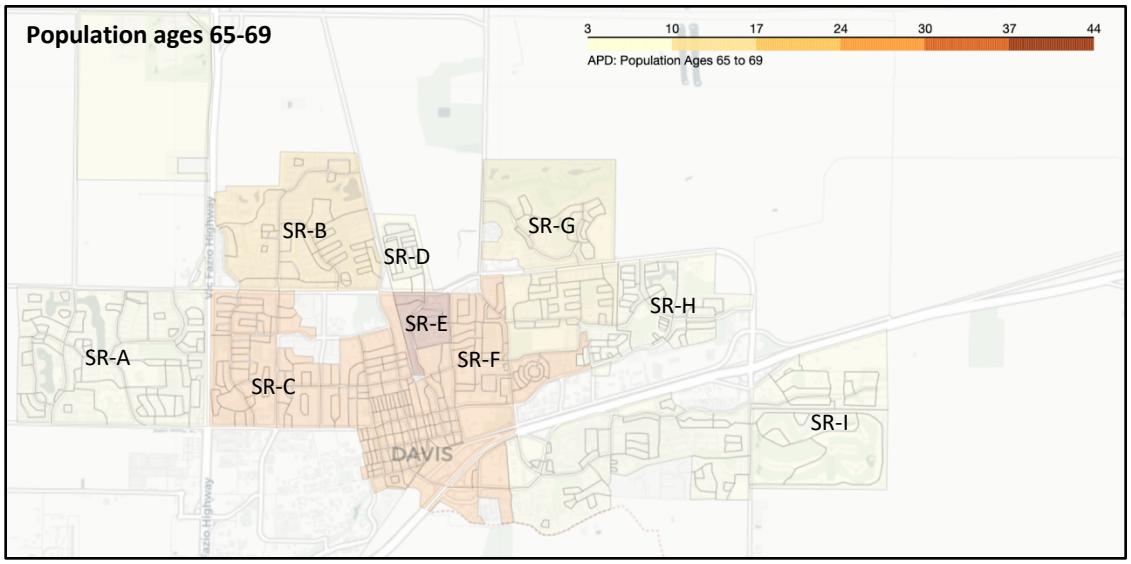
e)


Population ages 50-59
0 5 11 16 21 27 32
APD: Population Ages 50 to 59
SR-B
SR-D
SR-G
SR-H
SR-E
SR-F
SR-A
SR-C
SR-I
DAVIS

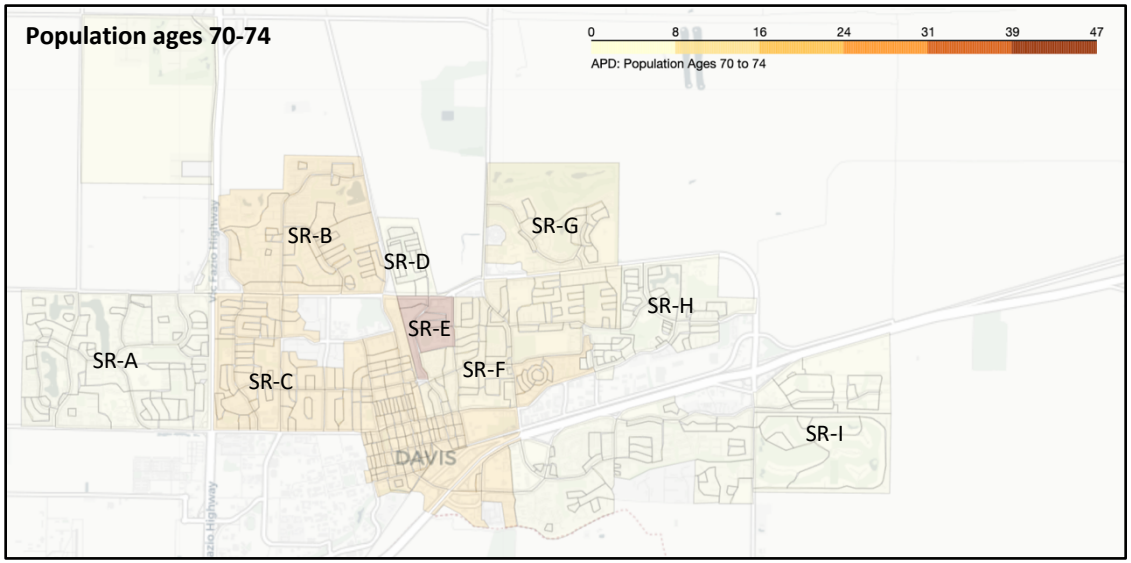
f)


Population ages 60-64
1 8 15 22 29 36 43
APD: Population Ages 60 to 64
SR-B
SR-D
SR-G
SR-H
SR-E
SR-F
SR-A
SR-C
SR-I
DAVIS

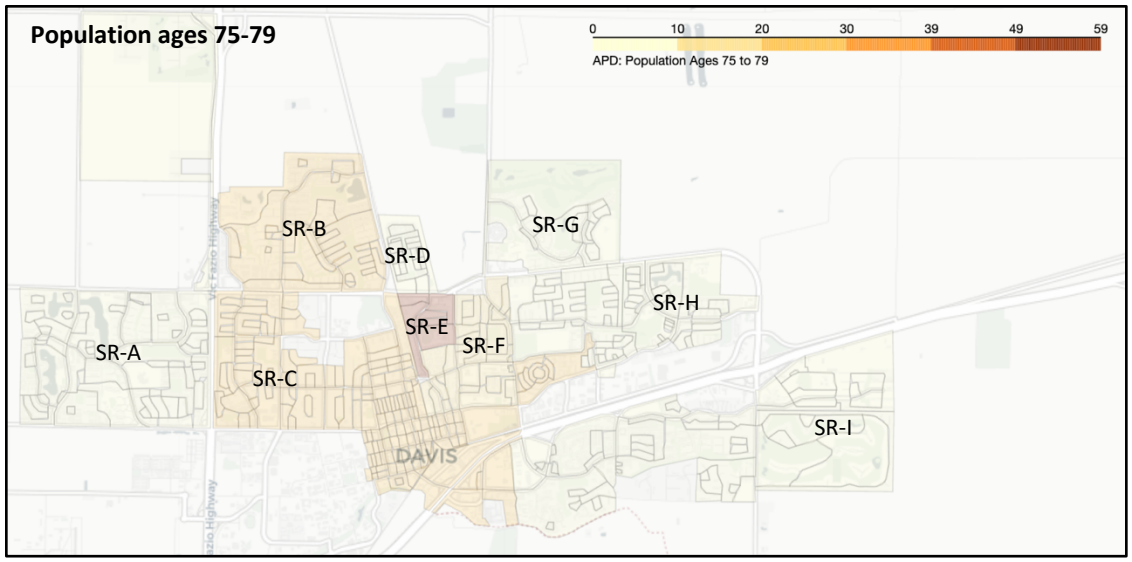
g)


Population ages 65-69
3 10 17 24 30 37 44
APD: Population Ages 65 to 69
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

h)


Population ages 70-74
0 8 16 24 31 39 47
APD: Population Ages 70 to 74
SR-B
SR-D
SR-G
SR-H
SR-E
SR-A
SR-F
SR-C
SR-I
DAVIS

i)


Population ages 75-79
0
10
20
30
39
49
59
APD: Population Ages 75 to 79
SR-B
SR-D
SR-G
SR-H
SR-E
SR-F
SR-A
SR-C
SR-I
DAVIS

j)

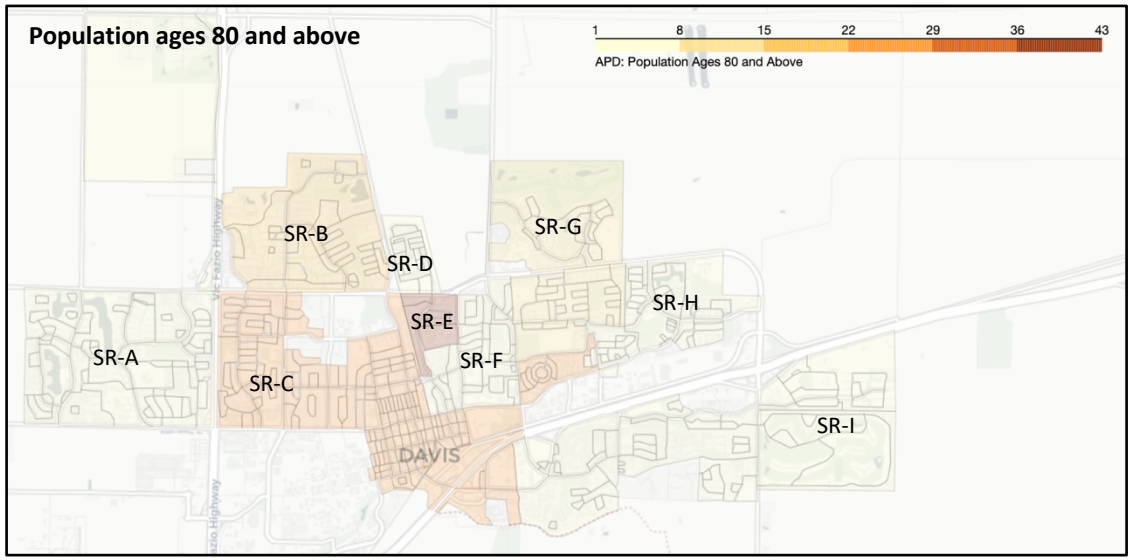


Figure 7. Choropleth maps of absolute percent differences (APDs) between probabilistic and manual assignment of census block data for age into sub-sewershed zones. Results are shown for each age category subgroup population captured through HDT wastewater sampling, by sub-sewershed zone. Stratifications included populations (a) under age 5, (b) ages 5 to 17, (c) ages 18 to 34, (d) ages 35 to 49, (e) ages 50 to 59, (f) ages 60 to 64, (g) ages 65 to 69, (h) ages 70 to 74, (i) ages 75 to 79, and (j) ages 80 and above.

Table 10. Absolute percent differences (APDs) between probabilistic and manual assignment of census block data for age into sub-sewershed zones.

Zone Name	Under 5	5 to 17	18 to 34	35 to 49	50 to 59	60 to 64	65 to 69	70 to 74	75 to 79	80 and Over
SR-A	5.1	4.3	6.5	4.0	4.8	3.7	8.4	3.9	3.4	1.8
SR-B	23.9	26.7	36.3	23.4	22.2	25.3	19.8	20.9	20.8	19.7
SR-C	24.7	25.3	21.5	29.7	26.9	26.6	31.1	24.5	25.7	25.6
SR-D	2.3	2.8	3.7	3.0	0.0	0.0	3.8	0.0	0.0	3.4
SR-E	34.0	36.3	27.0	32.6	32.8	42.9	45.0	47.1	60.0	43.9
SR-F	19.2	14.8	18.1	13.6	8.6	9.5	23.2	13.5	12.3	4.3
SR-G	15.8	16.2	20.2	18.0	20.4	14.8	12.9	11.7	9.7	13.8
SR-H	5.3	9.1	5.4	4.7	7.4	7.7	4.5	8.2	3.9	4.8
SR-I	5.3	5.4	2.7	3.4	3.3	4.1	3.4	3.0	1.9	23.2

REFERENCES

Bailit, M., & Kanneganti, D. (n.d.). A Typology For Health Equity Measures. *Health Affairs Forefront*. https://doi.org/10.1377/forefront.20220318.155498

Borchardt, M. A., Boehm, A. B., Salit, M., Spencer, S. K., Wigginton, K. R., & Noble, R. T. (2021). The Environmental Microbiology Minimum Information (EMMI) Guidelines: qPCR and dPCR Quality and Reporting for Environmental Microbiology. *Environmental Science & Technology*, *55*(15), 10210–10223. https://doi.org/10.1021/acs.est.1c01767

Bureau, U. C. (n.d.-a). *American Community Survey 5-Year Data (2009-2022)*. Census.Gov. Retrieved February 19, 2024, from https://www.census.gov/data/developers/data-sets/acs-5year.html

Bureau, U. C. (n.d.-b). *Increased Margins of Error in the 5-Year Estimates Containing Data Collected in 2020*. Census.Gov. Retrieved November 20, 2023, from https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes/2022-04.html

Bureau, U. C. (n.d.-c). *Understanding Geographic Identifiers (GEOIDs)*. Census.Gov. Retrieved January 11, 2024, from https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html

Bureau, U. C. (n.d.-d). *What are census blocks?* Census.Gov. Retrieved January 29, 2024, from https://www.census.gov/newsroom/blogs/random-samplings/2011/07/what-are-census-blocks.html

CDC. (2020, February 11). *Healthcare Workers*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html

CDC. (2023a, March 14). *National Wastewater Surveillance System*. Centers for Disease Control and Prevention. https://www.cdc.gov/nwss/wastewater-surveillance.html

CDC. (2023b, March 15). *CDC Museum COVID-19 Timeline*. Centers for Disease Control and Prevention. https://www.cdc.gov/museum/timeline/covid19.html

Census Bureau Data. (n.d.). Retrieved November 3, 2023, from https://data.census.gov/

Ceres—Protocols. (n.d.). Ceres Nano. Retrieved February 7, 2024, from https://www.ceresnano.com/protocols

Cheng, K. J. G., Sun, Y., & Monnat, S. M. (2020). COVID-19 Death Rates Are Higher in Rural Counties With Larger Shares of Blacks and Hispanics. *The Journal of Rural Health*, *36*(4), 602–608. https://doi.org/10.1111/jrh.12511

COVID-19 Age, Race and Ethnicity Data. (n.d.). Retrieved November 14, 2023, from https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/Age-Race-Ethnicity.aspx

COVID-19 Death Data and Resources—National Vital Statistics System. (2023, June 23). https://www.cdc.gov/nchs/nvss/covid-19.htm

Daza-Torres, M. L., Montesinos-López, J. C., Kim, M., Olson, R., Bess, C. W., Rueda, L., Susa, M., Tucker, L., García, Y. E., Schmidt, A. J., Naughton, C. C., Pollock, B. H., Shapiro, K., Nuño, M., & Bischel, H. N. (2023). Model training periods impact estimation of COVID-19 incidence from wastewater viral loads. *Science of The Total Environment*, *858*, 159680. https://doi.org/10.1016/j.scitotenv.2022.159680

Disclosure Avoidance and the 2020 Census: How the TopDown Algorithm Works. (n.d.).

Guan, A., Thomas, M., Vittinghoff, E., Bowleg, L., Mangurian, C., & Wesson, P. (2021). An investigation of quantitative methods for assessing intersectionality in health research: A systematic review. *SSM - Population Health*, *16*, 100977–100977. https://doi.org/10.1016/j.ssmph.2021.100977

Heller, J., Givens, M., Yuen, T., Gould, S., Benkhalti, M., Bourcier, E., & Choi, T. (2014). Advancing Efforts to Achieve Health Equity: Equity Metrics for Health Impact Assessment Practice. *International Journal of Environmental Research and Public Health*, *11*, 11054–11064. https://doi.org/10.3390/ijerph111111054

Holm, R. H., Osborne Jelks, N., Schneider, R., & Smith, T. (2023). Beyond COVID-19: Designing Inclusive Public Health Surveillance by Including Wastewater Monitoring. *Health Equity*, *7*(1), 377–379. https://doi.org/10.1089/heq.2022.0055

Holm, R. H., Pocock, G., Severson, M. A., Huber, V. C., Smith, T., & McFadden, L. M. (2023). Using wastewater to overcome health disparities among rural residents. *Geoforum; Journal of Physical, Human, and Regional Geosciences*, *144*, 103816. https://doi.org/10.1016/j.geoforum.2023.103816

Hu, X. C., Reckling, S. K., & Keshaviah, A. (2023). *Assessing health equity in wastewater monitoring programs: Differences in the demographics and social vulnerability of sewered and unsewered populations across North Carolina* [Preprint]. Public and Global Health. https://doi.org/10.1101/2023.10.06.23296680

Introduction to Digital PCR | Bio-Rad. (n.d.). Retrieved October 24, 2023, from https://www.bio-rad.com/en-us/life-science/learning-center/introduction-to-digital-pcr

Juel, M. A. I., Stark, N., Nicolosi, B., Lontai, J., Lambirth, K., Schlueter, J., Gibas, C., & Munir, M. (2021). Performance evaluation of virus concentration methods for implementing SARS-CoV-2 wastewater based epidemiology emphasizing quick data turnaround. *The Science of the Total Environment*, *801*, 149656. https://doi.org/10.1016/j.scitotenv.2021.149656

Kadonsky, K., Naughton, C., Susa, M., Olson, R., Singh, G., Daza-Torres, M., Montesinos-López, J., Garcia, Y., Gafurova, M., Gushgari, A., Cosgrove, J., White, B., Boehm, A., Wolfe, M., Nuño, M., & Bischel, H. (2023). *Expansion of wastewater-based disease surveillance to improve health equity in Californias Central Valley: Sequential shifts in case-to-wastewater and hospitalization-to-wastewater ratios.* https://doi.org/10.3389/fpubh.2023.1141097

Li, C., Bayati, M., Hsu, S.-Y., Hsieh, H.-Y., Lindsi, W., Belenchia, A., Zemmer, S. A., Klutts, J., Samuelson, M., Reynolds, M., Semkiw, E., Johnson, H.-Y., Foley, T., Wieberg, C. G., Wenzel, J., Lyddon, T. D.,

LePique, M., Rushford, C., Salcedo, B., … Lin, C.-H. (2022). *Population Normalization in SARS-CoV-2 Wastewater-Based Epidemiology: Implications from Statewide Wastewater Monitoring in Missouri* (p. 2022.09.08.22279459). medRxiv. https://doi.org/10.1101/2022.09.08.22279459

Mandelbaum, J., & Link to external site, this link will open in a new tab. (2020). Advancing health equity by integrating intersectionality into epidemiological research: Applications and challenges. *Journal of Epidemiology and Community Health*, *74*(9), 761–762. https://doi.org/10.1136/jech-2020-213847

Medina, C. Y., Kadonsky, K. F., Roman, F. A., Tariqi, A. Q., Sinclair, R. G., D'Aoust, P. M., Delatolla, R., Bischel, H. N., & Naughton, C. C. (2022). *The need of an environmental justice approach for wastewater based epidemiology for rural and disadvantaged communities: A review in California*. https://escholarship.org/uc/item/1d29z5kp

National Academies of Sciences, Engineering, and Medicine. (2023). *Wastewater-based Disease Surveillance for Public Health Action* (p. 52). The National Academies Press.

Nuño, M. (2024, January 31). *Sampling Optimization Frameworks*. Healthy Central Valley Together Research Workshop.

Planning and Zoning | City of Davis, CA. (n.d.). Retrieved November 13, 2023, from https://www.cityofdavis.org/city-hall/community-development-and-sustainability/planning-and-zoning

Pollock, B. H., Bergheimer, C. L., Nesbitt, T. S., Stoltz, T., Belafsky, S. R., Burtis, K. C., Carey, K. M., & Nuño, M. (2022). Healthy Davis Together: Creating a Model for Community Control of COVID-19. *American Journal of Public Health*, *112*(8), 1142–1146. https://doi.org/10.2105/AJPH.2022.306880

Propagation of Error. (2013, October 2). Chemistry LibreTexts. https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Supplemental_Modules_(Analytical_Chemistry)/Quantifying_Nature/Significant_Digits/Propagation_of_Error

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

Safford, H., Zuniga-Montanez, R. E., Kim, M., Wu, X., Wei, L., Sharpnack, J., Shapiro, K., & Bischel, H. N. (2022). Wastewater-Based Epidemiology for COVID-19: Handling qPCR Nondetects and Comparing Spatially Granular Wastewater and Clinical Data Trends. *ACS ES&T Water*, *2*(11), 2114–2124. https://doi.org/10.1021/acsestwater.2c00053

Spearman Rank—An overview | ScienceDirect Topics. (n.d.). Retrieved January 29, 2024, from https://www.sciencedirect.com/topics/engineering/spearman-rank

U.S. Census Bureau QuickFacts: Davis city, California. (n.d.). Retrieved October 23, 2023, from https://www.census.gov/quickfacts/fact/table/daviscitycalifornia/PST045222