Full title: Assessment of sewer connectivity in the United States and its implications for equity in wastewater-based epidemiology

Short title: Sewer connectivity and equity in wastewater-based epidemiology

Authors: QinQin Yu[1], Scott W. Olesen[2], Claire Duvallet[2], Yonatan H. Grad[1*]

Affiliations:

1. Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA
2. Biobot Analytics, Inc., Cambridge, Massachusetts, USA

[*]Corresponding author

Email: ygrad@hsph.harvard.edu (YG)

Abstract

Wastewater-based epidemiology is a promising public health tool that can yield a more representative view of the population than case reporting. However, only about 80% of the U.S. population is connected to public sewers, and the characteristics of populations missed by wastewater-based epidemiology are unclear. To address this gap, we used publicly available datasets to assess sewer connectivity in the U.S. by location, demographic groups, and economic groups. Data from the U.S. Census' American Housing Survey revealed that sewer connectivity was lower than average for American Indian and Alaskan Native, White, non-Hispanic, older, not in poverty, and larger households, but smaller geographic scales revealed local variations from this national connectivity pattern. For example, data from the U.S. Environmental Protection Agency showed that sewer connectivity was positively correlated with income in Minnesota, Florida, and California. Data from the U.S. Census' American Community Survey and Environmental Protection Agency also revealed geographic areas with low sewer connectivity, such as Alaska, the Navajo Nation, Minnesota, Michigan, and Florida. However, with the exception of the U.S. Census data, there were inconsistencies across datasets. Using mathematical modeling to assess the impact of wastewater sampling inequities on inferences about epidemic trajectory at a local scale, we found that in some situations, even weak connections between communities may allow wastewater monitoring in one community to serve as a reliable proxy for an interacting community with no wastewater monitoring, when cases are widespread. A systematic, rigorous assessment of sewer connectivity will be important for ensuring an equitable and informed implementation of wastewater-based epidemiology as a public health monitoring system.

Introduction

Wastewater-based epidemiology (WBE) plays an important role in surveillance of SARS-CoV-2 [1,2], polio virus [3,4] and other pathogens [5,6] and has applications to monitoring a variety of other public health concerns [7], including opioid usage [8]. One proposed benefit of wastewater-based epidemiology is that wastewater data is more representative of the population than case reporting, which can be biased towards those with health-seeking behavior or access to healthcare. For example, the populations served by the Health and Human Services SARS-CoV-2 National Wastewater Surveillance System were more representative of the entire US's age distribution and Black and Hispanic populations than the vaccinated population [9].

While WBE offers convenient sampling of populations served by public sewers, about 20% of individuals in the US live in homes not connected to public sewers [10,11]. This includes those on decentralized wastewater systems and those with no wastewater treatment systems. The most common decentralized wastewater system is septic tanks [12], which collect and treat wastewater onsite, typically the yard of the home. Households without wastewater treatment systems may have outhouses or privies, chemical toilets, or no plumbing. Variability in sewer connectivity exists across the U.S., with more connectivity in urban areas than rural areas [10,13].

As with any emerging public health tools, it is important to ask to what extent the tool exacerbates or alleviates inequities. A few studies have evaluated equity of sewer connectivity on broad geographic scales. A 2017 study from Environmental Protection Agency Office of Water showed that households in the U.S. that earned less than the national median household income (MHI, $61,000) were almost 10% more likely to have a decentralized wastewater

system or no wastewater treatment system compared to households that earned more than the MHI [14]. Additionally, this study found that as household income decreased, decentralized wastewater system usage increased in Florida, Hawaii, and Delaware, but not in Rhode Island. Based on the 2019 U.S. Census Bureau's American Housing Survey, rural areas are less connected to sewers, and while the income in rural areas was lower than in metropolitan areas, households connected to septic tanks were wealthier than those connected to sewers in both rural and metropolitan areas [10].

However, we have an incomplete understanding of the factors associated with sewer connectivity across the US and the implications for the interpretation of wastewater data. In this study, we sought to address the following questions, focusing our analyses on the U.S.: (1) To what extent is there demographic and economic inequity in sewer connectivity? (2) Which geographic areas have low sewer connectivity? (3) What is the applicability of WBE data to neighboring unsampled communities? To address the first question, we analyzed household-level data of sewer connectivity stratified by geographic, demographic, and economic variables. For the second question, we evaluated datasets aggregated at the county or county subdivision levels to qualitatively identify geographic areas that have low sewer connectivity. To address the third question, we used a mathematical model to simulate WBE in two interacting populations.

Methods

Dataset compilation

We used a convenience sample of publicly available datasets on sewer connectivity, septic connectivity, or no plumbing at the household level. The types of data included household-level data, locations of sewer systems and population served, and population served by or lacking

wastewater systems aggregated by county or county subdivision. All datasets are summarized in **Table *1*** and described in more detail in the Supplementary Information.

Table 1: Datasets used in this study.

Dataset	Location	Year(s)	Description	Sampling method	Dataset completeness and potential biases	Ref.
American Housing Survey	US excluding Island areas	2013, 2019, 2021	This was a longitudinal survey of housing, demographic, and economic characteristics of approximately 60,000-80,000 representative households.	The sampling method was a representative sample of all U.S. households at the time of sample selection (most recently in 2015 and before that, in 1985). New housing units were added each survey cycle.	The survey excludes group quarters, businesses, hotels, and motels. Additionally, the survey was more likely to classify rural seasonal homes as vacant units, counted large cluster septic systems as public sewer, and was thought to lose new households built in rural areas due to its longitudinal design.	[11,15–19]
American Community Survey 5-year estimates	US excluding Island areas	2021	This was a monthly survey of characteristics of populations and households of a representative subsample of addresses in the US. The survey reports the number of households lacking complete plumbing facilities by geographic subdivision.	The survey was sent to a random sample of addresses in the U.S. every month. Each address was selected no more than once every 5 years.	The surveys were completed by self-response (although a subset of non-responses were followed up with personal visits). Responses were reviewed for completeness and questionnaires needing clarification were followed up by phone calls. The data was averaged over 5 years to reduce statistical fluctuations.	[20]
Florida Department of Health	Florida	Up to 2012	This dataset includes the onsite sewage treatment and disposal systems locations	The sample included onsite sewage treatment and disposal system permits recorded in the	As of 2022, a permit was required for construction and inspection was recommended every 3-5 years after. From this dataset, an estimated 7% of households	[22]

septic tank inspections			inspected by the Florida Department of Health reported in June 2012.	Florida Department of Health system.	were connected to septic tanks across the state. This was substantially less than the Florida Department of Environmental Protection's estimate that approximately one third of Florida's population used septic tanks [21].	
EPA Clean Watersheds Needs Survey	US	2012	This survey included voluntary submissions of the locations of publicly owned wastewater collection and treatment facilities and their estimated population served for the purposes to assess funding needs for treatment works projects.	The responses were coordinated by states.	This survey had voluntary responses and variable effort and resources that each state put into the survey. States that had the most comprehensive responses were New York, California, Florida, New Jersey, Maryland, Iowa, Minnesota and Michigan. Missing facilities included those in South Carolina, the Northern Mariana Islands, and American Samoa; facilities whose projects did not have documented solutions or cost estimates; privately owned wastewater facilities; facilities on tribal lands and Alaskan Native Villages; and facilities whose projects received funding from other sources. Small community facilities are thought to be underrepresented due to having less resources for completing the survey, but this was not quantified.	[23,24]
US Census Island Areas Decennial Survey	US Island Areas	2020	This was a survey of housing, social, and economic information of all housing units.	The survey was conducted via in person and phone interviews.	The COVID-19 pandemic impacted collection of data on group quarters [25] and the survey had higher nonresponse rates than previous years' surveys, particularly for Guam. Quality control checks were performed on responses.	[26]

The datasets are further described in the S1 Appendix and the dataset completeness and potential biases are further assessed in the S2 Appendix. Additional datasets that are briefly

described in the main text but primarily analyzed in the Supplementary information are summarized in **Table S *3***.

Calculation of weights, error bars, and summary statistics in American Housing Survey (AHS) data

All summary statistics were calculated by applying the weights reported in the AHS Public Use File [19]. Because not every household was sampled and there was uneven sampling of households across the U.S., the weight estimates the number of similar households that each surveyed household represented. Error bars on summary statistics represent the middle 95% of values calculated using the reported replicate weights (160 replicate weights per household). Each summary statistic was calculated for each replicate separately before taking the middle 95% of values. Only categories with at least 5 households were used.

Analysis of sewer connectivity by urban or rural areas in AHS data

We used the most recent AHS National Sample with publicly available data on the urban or rural location of the household (2013) which categorized households into being in a central city of a Metropolitan Statistical Area (MSA), in an urban area of an MSA but not in a central city, in a rural area of an MSA but not in a central city, in an urban outside of an MSA, or in a rural area outside of an MSA [27]. Central cities are defined as having either (1) ≥250,000 population or at least 100,000 people working within corporate limits, (2) ≥25,000 population, at least 75 jobs for each 100 residents who were employed, and 60% or fewer of the city's resident workers commuted to jobs outside, or (3) 15,000-25,000 population, at least one third the size of the metropolitan statistical area's largest city, and met the two commuting requirements in (2). Metropolitan Statistical Areas were defined as whole counties that have significant levels of

community and contiguous urban areas in common and includes rural areas. Urban areas were defined as having ≥2500 people, with at least 1500 residing outside of institutional group quarters. Rural areas were defined as those not in urban areas.

Analyses of geographic areas with low connectivity to sewer

Data of individual locations of septic tanks or sewer collection systems, when they existed, were aggregated based on latitude and longitude into county or county subdivision boundaries as defined by the U.S. Census' TIGER shapefiles of the corresponding year [28]. To calculate the fraction of a geographic area connected to septic or sewer, the total population size or total number of households in that geographic area was taken from the corresponding year in the American Community Survey (ACS) 5-year estimate [20]. For county subdivision and census designated place analyses, only areas with at least 5 households and 20 population size were used for the analyses. Detailed data analyses of the state and island area datasets are described in the Supplementary information.

Counties and county subdivisions were classified into Metropolitan statistical areas (at least one urbanized area of 50,000 or more population), Micropolitan statistical areas (at least one urban cluster of 10,000-50,000 population), or Rural areas (all other areas) using the U.S. Census classification of the encompassing Core Based Statistical Area (core area containing a substantial population nucleus together with adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties) [29].

Demographic and economic variable correlations with sewer or septic connectivity

To assess correlations of demographic and economic variables with sewer and septic connectivity when aggregating at the county, county subdivision, or census designated place levels, we used the ACS 5-year estimates from the corresponding year and geographic scale. Only geographic subdivisions with at least 5 households and 20 population size were used. The Pearson correlation coefficient between demographic or economic variables of interest and the percentage of households in a geographic subdivision connected to septic tanks or sewer was calculated using the stats.pearsonr function in the scipy package (version 1.10.1) in Python (version 3.11.0). A Bonferroni correction was applied to correct for multiple hypothesis testing to calculate the q values.

Simulations of interacting populations and wastewater sampling

Two interacting populations (A and B) were modeled in a deterministic compartmental model with susceptible, infected, and recovered (SIR) compartments for each population and a fraction, ε, of cross-population contacts. The base parameters were set at population size N_A=N_B=5000, recovery rate γ_I=0.18 inverse days (corresponding to an infectious period of 5.6 days), basic reproduction number $R_0^A = R_0^B = \frac{\beta_A}{\gamma_I} = \frac{\beta_B}{\gamma_I} = 1.5$ and were varied in sensitivity analyses. The differential equations were solved using the integrate.odeint function in the scipy package (version 1.6.2) in Python (version 3.11.0) using a timestep of 0.1 days.

Wastewater sampling was modeled by the number of copies of pathogen genetic material shed per day by each population divided the volume of sampled wastewater produced per day by each population. Both numerator and denominator depended on the fraction of waste produced

(including shed pathogen genetic material) by each population that is sampled by wastewater, f_A and f_B. A detailed description of the simulations is provided in S1 Appendix.

Code availability

All code and data to reproduce the analyses and simulations are available at https://github.com/gradlab/wastewater_equity.

Results

Is there inequity in who is sampled by wastewater?

To address this question, we used the American Housing Survey (AHS), a biannual longitudinal survey of representative housing units within the U.S that reports household-level data of sewer connectivity, demographic factors, and economic factors (**Table *1***). We focused our analyses on geographic scales with a reliable sample size for statistical analyses (census division, urban/rural at the national level, metropolitan statistical areas) [30].

According to the 2021 AHS, 83% of US households were connected to public sewers and 16% to standard septic tanks (**Fig S *1***). Each of the other forms of sewage disposal (non-standard septic tanks, chemical toilets, outhouses or privies, other, none, or not reported) was more than an order of magnitude less common (**Fig S *1***). Across census divisions in 2021, the western US was overall better connected to sewers than the eastern US, with the highest levels of sewer connection in the Pacific (91.3% of households, CI: 90.8%-91.7%) and Mountain (87.6% of households, CI: 86.8%-88.4%) divisions and the lowest levels of sewer connection in the New England (72.9% of households, CI:72.1%-74.0%) and East South Central (74.9% of households, CI: 73.7%-76.3%) divisions (**Fig *1***).

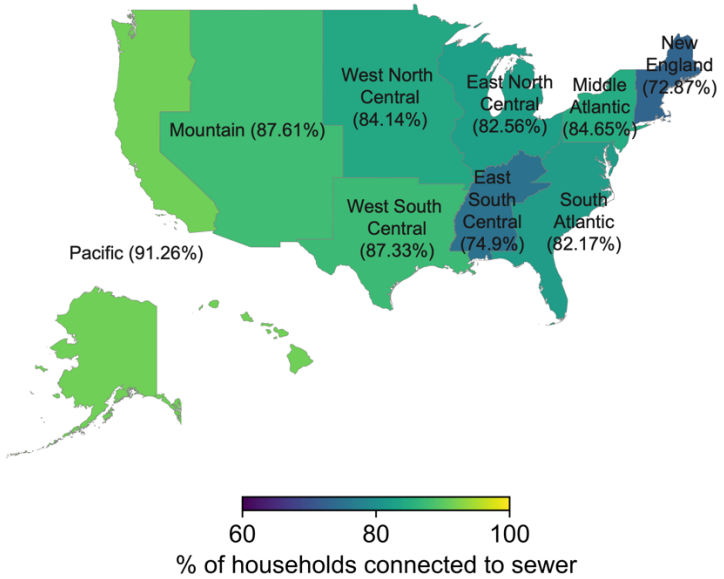


Fig 1*:* The percentage of households connected to sewers in each census division. Data are from the 2021 American Housing Survey National Survey [11].

Across census divisions, households with an Asian, Black or African American, or Native Hawaiian and Other Pacific Islander householder (the owner or renter of the unit) were on average 14.0%, 13.1%, and 12.5% more connected to sewers, respectively, than the overall census division; however, specifically in the Pacific census division, households with a Native Hawaiian and Other Pacific Islander householder were 5.3% less connected to sewers than the overall census division (**Fig 2**). Households with an American Indian and Alaska Native or White householder were both on average 2.6% less connected to sewers than the overall census division. Across census divisions, households with a Hispanic householder were on average 11.7% more connected to sewers whereas households with a non-Hispanic householder were 1.1% less connected to sewers than the overall census division.

Connection to sewers decreased with the age of the householder (**Fig 2**), with the exception that households with a householder aged 75+ were better connected in most census divisions than those with a householder aged 65-74 (79.7% vs 77.9% of households). Households with 1 person were on average 4.2% more connected whereas households with 2 people were on average 3.6% less connected to sewers than the overall census division. There were no consistent trends across census divisions for households with more than 2 people. Additionally, households that were in poverty were better connected to sewers (average of 87.5% of households across census divisions) than households not in poverty (82.7% of households across census divisions; **Fig 2**). The threshold used to determine whether a household was in poverty depended on family size and age of persons in the household; for example, in 2021 the threshold for 1 person under 65 years is $14,097 and for 4 persons with two children is $27,479 [31].

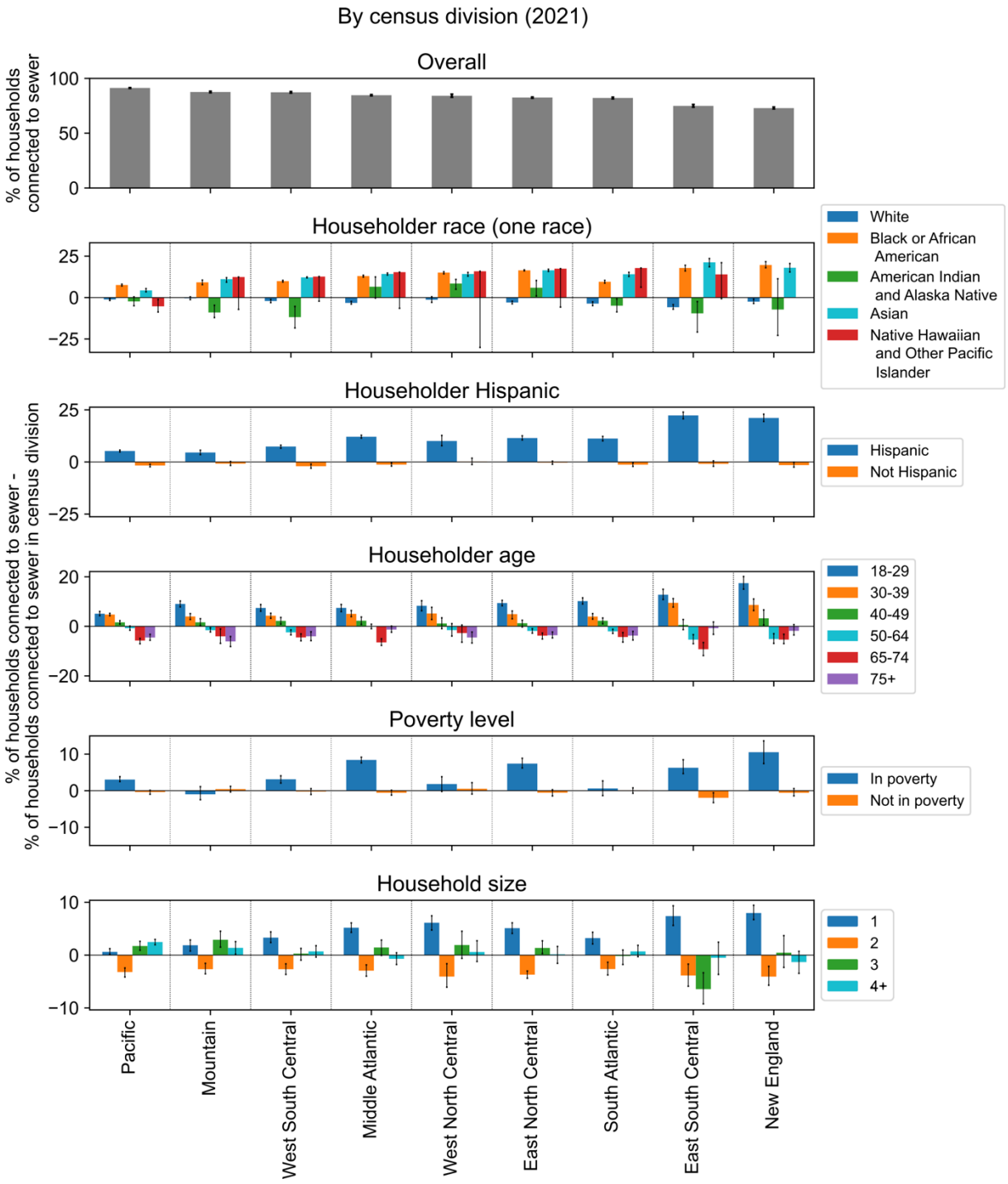
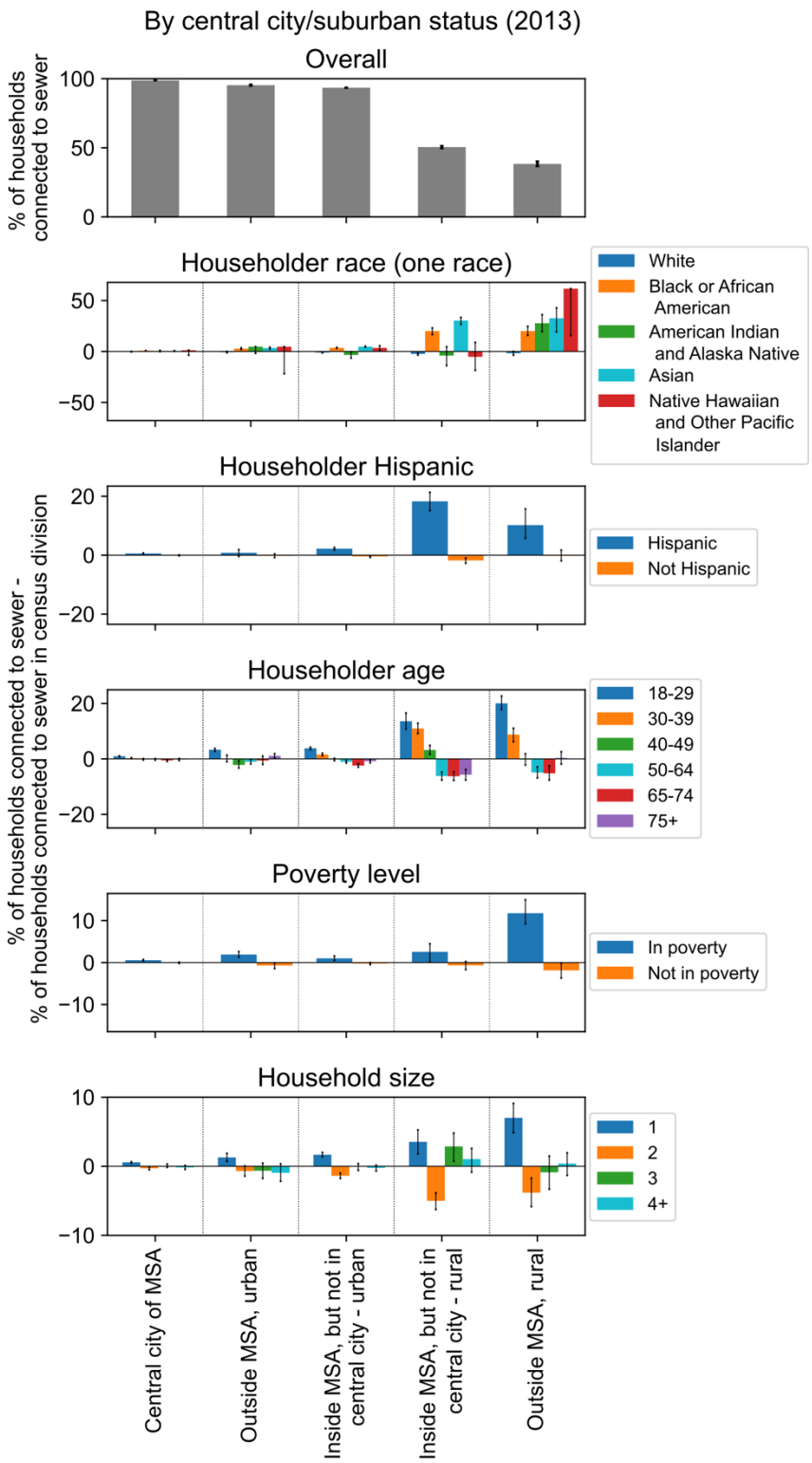


Fig 2*:* The percentage of households connected to sewers in each census division by demographic and economic characteristics. Data are from the 2021 American Housing Survey National Survey [11].

We sought to describe how the differences in sewer connection between demographic and economic groups observed at the census division level was affected by household location in urban or rural areas (see Methods). The qualitative trends observed at the census division level for race, ethnicity, age, household size, and poverty level were largely preserved across national urban and rural areas (**Fig 3**). One notable exception was that lower sewer connectivity for households with an American Indian and Alaska Native or Hawaiian and Other Pacific Islander householder compared to those with a Black or Asian householder was only observed in locations inside of an Metropolitan Statistical Area (MSA) but outside of a central city (both urban and rural). In central cities and locations outside of an MSA (both urban and rural), households with an American Indian and Alaska Native or Hawaiian and Other Pacific Islander householder had similar levels of sewer connectivity as households with a Black or Asian householder.


By central city/suburban status (2013)
Overall
% of households connected to sewer
100
50
0
Householder race (one race)
50
0
−50
White
Black or African American
American Indian and Alaska Native
Asian
Native Hawaiian and Other Pacific Islander
Householder Hispanic
20
0
−20
Hispanic
Not Hispanic
% of households connected to sewer - % of households connected to sewer in census division
Householder age
20
0
−20
18-29
30-39
40-49
50-64
65-74
75+
Poverty level
10
0
−10
In poverty
Not in poverty
Household size
10
0
−10
1
2
3
4+
Central city of MSA
Outside MSA, urban
Inside MSA, but not in central city - urban
Inside MSA, but not in central city - rural
Outside MSA, rural

231

Fig 3*:* The percentage of households connected to sewers by urban and rural categorizations and by demographic and economic characteristics. See Methods for definitions of Metropolitan Statistical Area (MSA), central city, urban, and rural. Data from the 2013 American Housing Survey National Survey [32]. MSA: metropolitan statistical area.

Next, we assessed sewer connectivity in large metropolitan areas, which was the smallest geographic scale resolved by the AHS due to sample size. In 35 of the largest metropolitan areas in 2019 and 2021, household connectivity to sewers ranges from 69.2% (Birmingham-Hoover, AL) to 99.4% (LA-Long Beach-Anaheim, CA) (**Fig S *2***). In each metropolitan area, households with a White or American Indian and Alaska Native householder, with an older householder, not in poverty, with a larger household size, and with a non-Hispanic householder were less connected to sewers than the average household, which was similar to what was seen when stratifying by census division. Exceptions were lower connectivity for households with a Hispanic householder in San Jose, CA (Hispanic householder: 96.1% of households, CI: 95.1%-97.2%; non-Hispanic householder: 97.8% of households, CI: 97.5%-98.2%) and lower connectivity for households in poverty in San Francisco, CA (in poverty: 97.7%, CI: 96.7%-98.7%; not in poverty: 99.3%, CI:99.1%-99.5%); Dallas, TX (in poverty: 92.5%, CI: 90.5%-94.5%; not in poverty: 95.1%, CI: 94.7%-95.6%); Chicago, IL (in poverty: 92.8%, CI:91.2%-94.3%, not in poverty: 94.8%, CI: 94.3%-95.3%); and Tampa, FL (in poverty: 83.9%, CI: 81.2%-87.0%; not in poverty: 88.8%, CI: 88.1%-89.6%) (**Fig S *2***).

In summary, at the census division level, broad trends of sewer connectivity by racial group revealed higher connectivity for households with an Asian, Black or African American, Native Hawaiian and Other Pacific Islander, or Hispanic householder and lower connectivity for households with an American Indian and Alaska Native or White householder. Sewer connectivity decreased with age, was highest in households with one person, and higher for

258 households in poverty. These trends were similar when stratifying by urban, rural, central city,
259 and MSA status, except that the broad racial disparities described above were only observed in
260 urban and rural areas that are inside of an MSA but outside of a central city. Broad trends were
261 also similar for individual large metropolitan areas, with some exceptions for households with a
262 Hispanic householders and households in poverty in particular MSAs.
263

264 Which geographic areas have low connectivity to sewers and 265 what are their demographics?

266 Next, we sought to assess which geographic areas in the U.S. have low connectivity to sewer
267 systems. These are areas where wastewater-based epidemiology may provide little direct
268 information on disease dynamics. Additionally, we assessed correlations of sewer connectivity
269 at the county or county subdivision level with demographic and economic factors. The
270 compilation of the datasets and an assessment of data completeness are described in the S2
271 Appendix, **Table *1***, and **Table S *3***.
272
273 Using data from the 2021 American Community Survey on the percentage of occupied housing
274 units lacking complete plumbing facilities by county, we defined the lower bound on the
275 percentage of the population not connected to sewers. In 14 counties, ≥10% of occupied
276 housing units lacked complete plumbing facilities; of these counties, 12 were in Alaska and 2
277 were in the Navajo Nation. In 5 counties, ≥20% of occupied housing units lacked complete
278 plumbing facilities, all in Alaska. The Yukon-Koyukuk Census Area in Alaska, had the highest
279 percentage of occupied housing units lacking complete plumbing facilities at 36% (**Fig S 3**,
280 **Table S *1***).
281

282 The Florida Department of Health onsite sewage treatment and disposal systems inspection
283 data (**Fig *4***) revealed that the percentage of households not on septic tanks (expected to
284 correlate with the percentage of households connected to sewer) was higher in metropolitan
285 compared to micropolitan county subdivisions and higher in micropolitan compared to rural
286 county subdivisions, but the effect was not significant ($p = 0.052$ and $p = 0.37$, respectively) (**Fig**
287 **S 4**a-b). The percentage of households not on septic tanks was significantly lower in county
288 subdivisions in the panhandle (74.0% vs 92.0%, $p<10^{-9}$) and in county subdivisions that did not
289 border the coast (86.6% vs 95.7%, $p<10^{-9}$) (**Fig S 4**). By combining this dataset with the 2012
290 American Community Survey (ACS) describing demographic and economic characteristics of
291 county subdivisions, we found that county subdivisions with less septic usage (suggesting more
292 sewer connectivity) were significantly more Hispanic, more Asian, and had a smaller household
293 size (**Table *2*** and **Fig S 5**), consistent with the nation-wide results from the AHS. These trends
294 were driven by metropolitan county subdivisions (**Fig S 6**), and an additional trend in
295 metropolitan county subdivisions not observed for all county subdivisions was that county
296 subdivisions with less septic usage had a significantly lower percentage of American Indian and
297 Alaska Natives. There were too few micropolitan and rural county subdivisions to observe
298 significant trends (**Fig S 7** and **Fig S *8***).
299

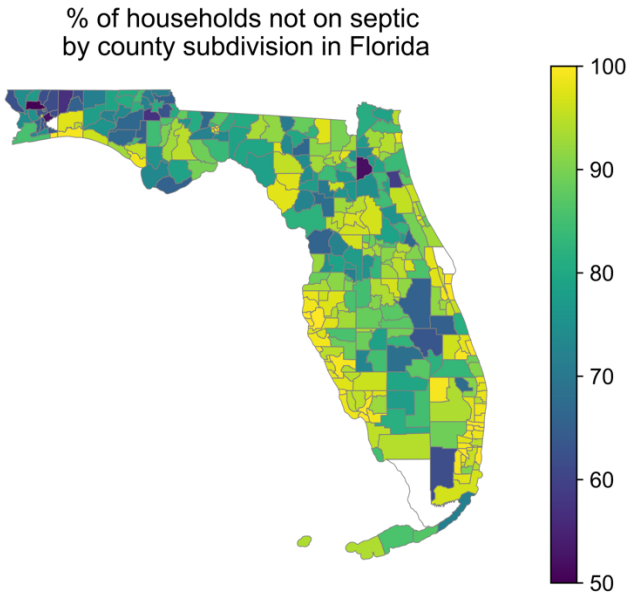


Fig 4*:* Map of the percentage of households in each county subdivision of Florida *not* connected to septic tanks. Data are from the Florida Department of Health septic tank inspection data from June 2012. County subdivisions with fewer than 5 households or 20 population size are displayed in white.

Table 2: Correlation of the percentage of a Florida county subdivision *not* connected to septic tanks with different demographic or economic variables.

Demographic or economic variable	Correlation of variable with % of county subdivision *not* on septic (Pearson correlation coefficient)	Significance (q-value)
Percent one race and Asian	0.31	**3.2E-07**
Percent Hispanic	0.21	**1.7E-03**
Average household size	-0.19	**9.1E-3**
Percent one race and American Indian and Alaska native	-0.15	0.07

Median age	0.12	0.36
Percent one race and White	-0.12	0.40
Percent one race and Black or African American	0.09	1.11
Median income	0.07	2.27
Percent one race and some other race	0.07	2.33
Percent one race and Native Hawaiian and other Pacific Islander	-0.04	5.76
Percent uninsured (health insurance)	-0.02	8.24

Correlations significant to 5% are shown in bold. Data are from the Florida Department of Health septic tank inspection permits reported in 2012.

In the 8 states with robust data in the EPA Clean Watersheds Needs Survey, multiple neighboring counties in Florida, Michigan, and Minnesota had less than 20% of residents receiving sewage collection, whereas this was not the case in California, Florida, Iowa, Maryland, New Jersey, and New York (**Fig S 9**). Across multiple states, sewer connectivity by county was positively correlated with percent Asian and percent Black or African American and negatively correlated with percent White (**Table 3**), consistent with the nationwide trends from the AHS data. While counties more connected to sewers appeared to have a significantly lower percent of American Indian and Alaskan Natives in California, this is affected by missing data from Indian reservations. Age and percent without health insurance were negatively correlated with sewer connectivity. Counties more connected to sewers had significantly higher income in Minnesota, Florida, and California, in contrast to the national results from the AHS data. Significant trends of sewer connectivity with demographic and economic factors were predominantly driven by metropolitan counties (**Table S *2***).

Table 3: Correlation of the percentage of households in a county connected to public sewers with different demographic or economic factors by state.

Demographic or economic variable	State	Correlation of variable with % of households by county connected to public sewers (Pearson correlation coefficient)	Significance (q-value)
Median age	MN	-0.43	2.9E-03
	MI	-0.41	1.1E-02
Median income	MN	0.44	1.7E-03
	FL	0.48	3.4E-03
	CA	0.49	7.2E-03
Percent one race and American Indian and Alaska Native	CA	-0.50	4.9E-03
Percent one race and Asian	MN	0.58	2.7E-07
	MI	0.51	7.7E-05
	NY	0.55	2.8E-04
	CA	0.56	3.9E-04
	FL	0.53	4.2E-04
Percent one race and Black or African American	MN	0.62	1.2E-08
	MI	0.59	4.6E-07
	NY	0.55	3.0E-04
	IA	0.38	7.7E-03
	CA	0.46	2.1E-02
	MI	-0.57	1.4E-06

	CA	-0.54	9.6E-04
Percent one race and White	NY	-0.51	1.8E-03
	IA	-0.35	2.8E-02
Percent uninsured (health insurance)	MI	-0.38	3.3E-02

Only correlations that are significant (q-value $\leq$ 0.05) are shown. Data are from the 2012 EPA Clean Watersheds Needs Survey.

Analysis of the Island Areas of Guam, the Northern Mariana Islands, the Virgin Islands, and American Samoa revealed lower levels of sewer connectivity than in the states, with considerable spatial variability in connectivity (see S2 Appendix).

Factors influencing applicability of wastewater-based epidemiology to communities lacking sewer connection

The inequities in sewer connections among communities raise the question of the applicability of inferences drawn from wastewater data in one community to communities lacking sewer connections. To explore this question, we used a deterministic compartmental model of two interacting populations with tunable levels of interaction, each with susceptible-infected-recovered (SIR) dynamics, and with wastewater sampling in each population into a common sample (**Fig *5***a). We set population A to be entirely connected to sewers and population B to have a tunable level of sampling by wastewater.

A common application of wastewater data is to aid in determining when an outbreak has peaked, which can inform policy decisions on when to ease restrictions. We first asked how well the wastewater data could predict the time of peak infections in population B when it is

completely unconnected to sewer systems and thus the only sampling is from population A. We found that when the two populations have similar sizes, the wastewater concentration peak and the infection peak in population B was within one generation time except when the interactions between the two population is weak (**Fig *5***b). The generation time is the average time between an individual's infection and transmission, which in an SIR model is the same as the infectious period. For example, in two populations with 5000 individuals each, a disease basic reproduction number of 1.5, and infectious period of 5.6 days, the percentage of contacts that occur across populations must drop below 4% before the peak in wastewater concentration occurred more than one generation time apart from the peak in infections in the unconnected population (**Fig *5***c).

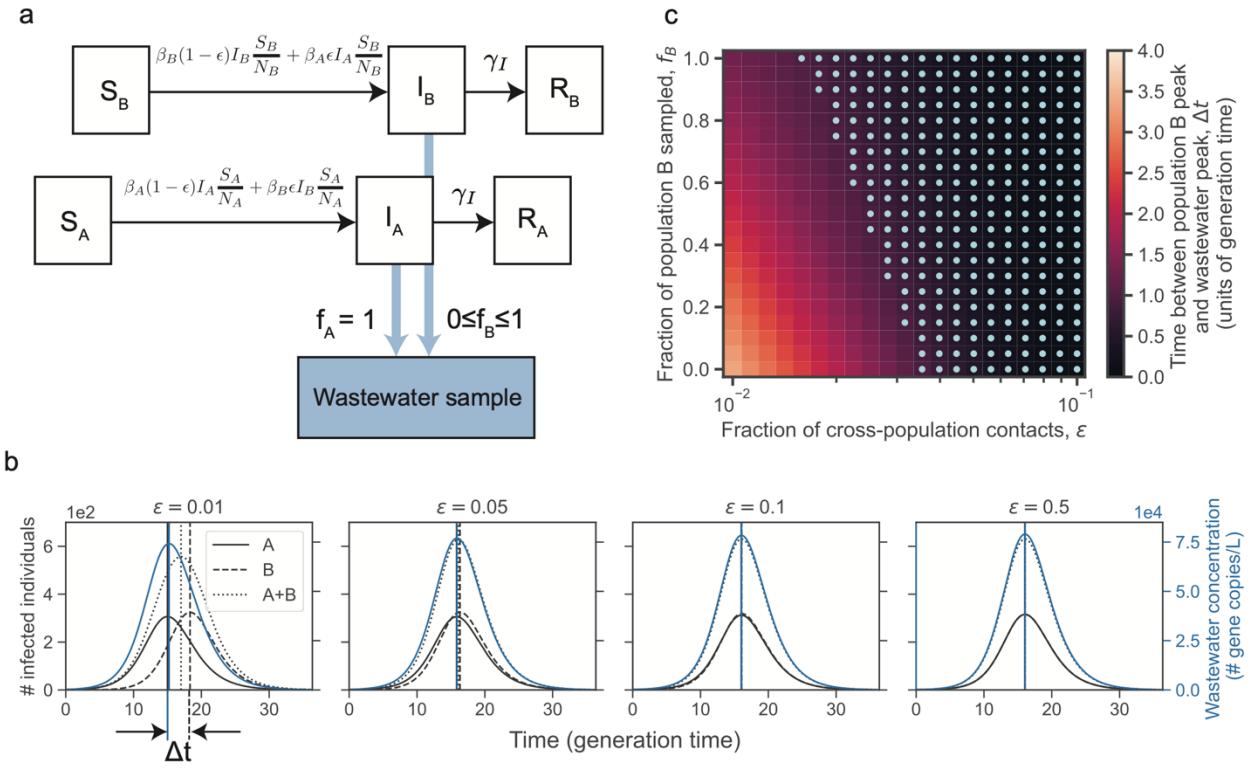


Fig 5: Deterministic compartmental model of two interacting populations with susceptible, infected, and recovered compartments and sampling by wastewater. (a) Schematic of model. Population A is entirely sampled by wastewater and a fraction of population B, f_B, is sampled by wastewater. (b) Number of infected individuals (black curves) in

population A, population B, and population A and B combined and wastewater concentration (blue curves) over time for different interaction strengths or fractions of cross-population contacts (ε). The vertical lines indicate the peak time for the curves with matching line style and color. The time between the population B peak in infections and the wastewater concentration peak is indicated by Δt. (c) The time between the population B peak in infections and the wastewater concentration peak (defined in (b)) when varying the fraction of cross-population contacts (ε) and the fraction of population B that is sampled (f_B). Light blue dots indicate parameter regimes in which Δt is less than 1 generation time. In both (b) and (c), the simulation starts with a single infected individual in population A at time 0 and no infected individuals in population B. The parameters are set to population size $N_A = N_B = 5000$, basic reproduction number $R^A{}_0 = R^B{}_0 = R_0 = \frac{\beta_A}{\gamma_I} = \frac{\beta_B}{\gamma_I} = 1.5$, rate of recovery $\gamma_I = 0.18$, and the overall contact rate times the probability of infection given contact $\beta_A = \beta_A = \gamma_I R_0$.

We then weakened the assumption that population B had no contribution to the sampled wastewater and allowed a fraction of population B to be sampled by wastewater. For example, individuals in population B may commute to a workplace sampled by wastewater, or a fraction of population B may be connected at home. As the fraction of population B that is sampled by wastewater increases, the time between infection and wastewater peaks decreases; however, the fraction of B that is sampled has a much weaker effect on the wastewater peak time than the interaction parameter between the communities. For example, with the above parameters and equal population sizes, with 10% sampling of population B, the cross-population contacts must drop below 4%, and with 50% below 3% and with 100% below 2% before the wastewater concentration peak and the peak in the number of infections in population B drops below one generation time.

Varying the population size of both populations (with an equal population size between the two populations), the recovery rate or generation time, and the R_0 of both populations (with an equal R_0 in the two populations) has only a weak effect on the discrepancy between wastewater concentration and population B infection peaks (**Fig S 10**).

When the unconnected population has a smaller size than the connected population, then even substantial interactions between the two populations can lead the wastewater concentration to peak before the infections in the unconnected population peaks (**Fig S 10**). Both very weak and very strong interactions cause the wastewater peak to coincide with the peak number of infections in the unconnected population.

Additionally, when the unconnected population has a smaller R_0 than the connected population (which could occur for instance if the unconnected population was more spread out geographically and thus had a lower contact rate), the outbreak in population B peaks later than the wastewater peak leading to a discrepancy in the peak times (**Fig S 10**).

Discussion

Equity of sewer connectivity

Our analysis of available datasets revealed considerable variability in sewer connectivity within and between communities, across locations, demographics, and economic statuses in the United States. The western US had higher levels of sewer connectivity than the eastern US. Across the US, lower connectivity to sewers was observed for American Indian and Alaska Native householders; White householders; non-Hispanic householders; older populations; and larger households in most census divisions. In the Pacific census division, lower connectivity to

sewers was observed for Native Hawaiian and Other Pacific Islander householders specifically in areas inside of a metropolitan statistical area but outside of a central city. Across census divisions, households not in poverty were less connected to sewers, consistent with the observation that, nation-wide, higher income households were less connected to sewers [10]. Aggregating by large metropolitan areas yielded similar trends as when aggregating by census divisions, except for lower connection for Hispanic householders in San Jose area and for households in poverty in San Francisco, Dallas, Chicago, and Tampa metropolitan areas, whereas these groups were better connected to sewers at the census division level. The differences in the MSA-level data suggested heterogeneity in connectivity across locations.

Geographic variation in sewer connectivity

Large parts of Alaska and the Navajo Nation lacked plumbing. This is consistent with the finding from the Annual Report to Congress on Sanitation Deficiency Levels for Indian Homes and Communities in 2019 that approximately 20% of Indian homes in Alaska do not have sewage disposal and about 10% of Indian homes in the Navajo Nation do not have sewage disposal [33].

State-level datasets revealed that sewer connectivity varied between and within communities. While in some states most areas had sewer connectivity (California, Iowa, Maryland, New Jersey, New York), in others there were large regions that lacked connectivity (Minnesota, Michigan, Florida). Asian and Hispanic populations were less connected to septic tanks across county subdivisions in Florida; however, there was considerable variability across the state and potential biases in the data. In contrast to the national data, in Minnesota, Florida, and California, counties with lower median household incomes were less connected to sewers, suggesting variability across the nation. This result in Florida is consistent with the observation

that as income levels rose, household decentralized system usage declined in Florida [14]. However, we did not observe any significant correlation of income with septic usage in the Florida Department of Health onsite sewage treatment and disposal systems inspection permitting dataset, possibly due to not all septic tanks having a permit that is reported in this dataset.

While these correlations may be largely driven by differences between urban and rural areas within a county or county subdivision, they may still be very useful in assessing inequities in sewer connectivity in large catchments that span neighboring counties or county subdivisions.

Data limitations

Substantial data gaps and biases prevented a comprehensive analysis of disparities in sewer system connectivity within and between communities across the US, particularly for small communities, tribal lands and Alaska Native Villages, and state and local geographic scales. The current design of the American Housing Survey has too small of sample size to study state and local geographic scales [30]. Additionally, the AHS may have underestimated the fraction of households using septic systems due to its survey design (**Table *1***). The differences between Florida state data supports the need for detailed and reproducible longitudinal studies of sewer connectivity (both in developing sampling and analysis methods).

Public health officials may find it beneficial to conduct local equity studies of sewer connectivity. Data on the catchment area sizes of the wastewater treatment plants would allow a better understanding of the geographic extent of the sampled population and the level of geographic aggregation needed to study sewer connectivity inequities. A recent study mapping wastewater treatment plant catchment areas and their population sizes served in New York state provides a

462 template for how to create these sewer catchment maps from a combination of permitting,
463 survey, and tax record data [13]. If possible, it would be helpful to consolidate and standardize
464 existing data where it exists (for instance, sewage disposal permitting at the local or state
465 levels). Additionally, if the EPA's request to expand the question on the ACS about access to
466 plumbing facilities to additionally ask about the type of plumbing facility [14,34] is granted, then
467 this would be a valuable dataset with the geographic and temporal resolution to study equity in
468 sewer connectivity moving forward.
469

470 Generalizability of wastewater-based epidemiology data in 471 light of heterogeneities in sewer connectivity

472 Our modeling results suggest that even weak interactions between two communities allow
473 wastewater monitoring in one community to serve as reliable proxy for the time of maximum
474 infections in the other community when the population sizes and R_0 of the two populations are
475 comparable, but not when the unconnected population has a substantially lower population size
476 or R_0. In the scenario with unequal population size, outbreaks that are seeded at the same time
477 in the two populations in the absence of interactions peak earlier in the unconnected population
478 due to its smaller size. With weak interactions, seeding of infections from population A to B
479 occurs slower, leading to more coincidence in peaks; with strong interactions, then the
480 dynamics of B are dominated by those in A, leading to more coincidence in peaks; at
481 intermediate interaction strengths, infections are seeded earlier and peaks early, leading to the
482 largest discrepancy in wastewater and infection peaks.
483
484 As the purpose of our model was to explore the impact of factors on the generalizability of
485 wastewater rather than to accurately capture the dynamics in all scenarios, we made simplifying
486 assumptions, including that all infected individuals shed the same amount of pathogen genetic

material into wastewater and shed only during the infected period; that all individuals contribute equally to wastewater; and that pathogen detection in wastewater has perfect sensitivity.

Our results suggest that in assessing the generalizability of wastewater data, it would be useful to estimate the extent of mobility between connected and unconnected communities. Interestingly, Ref [35] found no correlation between the size of a catchment area and the correlation of wastewater with case data for SARS-CoV-2, consistent with the result that interactions between connected and unconnected communities cause the disease dynamics to look similar to the wastewater data. In regions without sewer connectivity or with little interaction between connected and unconnected communities, wastewater data from neighboring communities will be less informative. In areas with low sewer connectivity in households, sampling wastewater outflow at frequently visited non-household locations (i.e., schools, offices, malls, etc.) may capture a more representative population.

Additional considerations for equity in WBE

In addition to inequities in sewer connectivity, the sewer locations used for wastewater sampling should be considered to promote demographic equity and ensuring the ability to capture spatiotemporal trends [36,37]. While we have focused on analysis on the US due to data availability, internationally disadvantaged populations are associated with lower access to sewers [38]. The ongoing development of wastewater sampling in non-sewered settings [39], for example in water channels in Las Vegas Valley [40], onsite sanitation facilities in Bangladesh, a refugee camp in Lebanon [41], and various non-sewered settings in low and middle income countries [42], represents a critical area for research and development.

Conclusions

In summary, while wastewater-based epidemiology is a useful tool to monitor disease burden and dynamics, our analyses suggest that access to this new tool varies across the US. More comprehensive data on sewer connectivity is needed, and in combination with assessments of mobility and population parameters, these data can help with the design of wastewater sampling schemes and the interpretation for epidemic trends in sampled and neighboring unsampled communities.

Acknowledgements

We thank Cara Omana from the Minnesota Pollution Control Agency for generously sharing the Minnesota WINS data and Harry Campbell from the Utah Department of Environment Quality, Division of Water Quality for generously sharing the Utah MWPP survey data and to both for answering questions about the datasets. We thank colleagues in the Grad lab and the Harvard T. H. Chan School of Public Health Center for Communicable Disease Dynamics for helpful discussions, particularly Stephen Kissler.

References

1. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat Biotechnol. 2020;38: 1164–1167.
2. 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. Curr Opin Environ Sci Health. 2020;17: 49–71.
3. Pöyry T, Stenvik M, Hovi T. Viruses in sewage waters during and after a poliomyelitis outbreak and subsequent nationwide oral poliovirus vaccination campaign in Finland. Appl Environ Microbiol. 1988;54: 371–374.
4. Link-Gelles R, Lutterloh E, Schnabel Ruppert P, Backenson PB, St George K, Rosenberg ES, et al. Public Health Response to a Case of Paralytic Poliomyelitis in an Unvaccinated Person and Detection of Poliovirus in Wastewater - New York, June-August 2022. MMWR Morb Mortal Wkly Rep. 2022;71: 1065–1068.
5. Kilaru P, Hill D, Anderson K, Collins MB, Green H, Kmush BL, et al. Wastewater Surveillance for Infectious Disease: A Systematic Review. Am J Epidemiol. 2023;192: 305–322.
6. Boehm AB, Hughes B, Duong D, Chan-Herur V, Buchman A, Wolfe MK, et al. Wastewater concentrations of human influenza, metapneumovirus, parainfluenza, respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids during the COVID-19 pandemic: a surveillance study. Lancet Microbe. 2023. doi:10.1016/S2666-5247(22)00386-X
7. Baz-Lomba JA, Salvatore S, Gracia-Lor E, Bade R, Castiglioni S, Castrignanò E, et al. Comparison of pharmaceutical, illicit drug, alcohol, nicotine and caffeine levels in wastewater with sale, seizure and consumption data for 8 European cities. BMC Public Health. 2016;16: 1035.

8. Gushgari AJ, Venkatesan AK, Chen J, Steele JC, Halden RU. Long-term tracking of opioid consumption in two United States cities using wastewater-based epidemiology approach. Water Res. 2019;161: 171–180.
9. Analytics B. Wastewater for equitable Covid-19 monitoring. In: Medium [Internet]. 8 Jul 2021 [cited 9 Feb 2023]. Available: https://biobotanalytics.medium.com/wastewater-for-equitable-covid-19-monitoring-e3b947d91e0a
10. Olesen SW, Young CCW, Duvallet C. White paper: The effect of septic systems on wastewater-based epidemiology. 2022. doi:10.5281/zenodo.7130726
11. U.S. Census Bureau. AHS 2021 National Public Use File. Available: https://www2.census.gov/programs-surveys/ahs/2021/AHS%202021%20National%20PUF%20v1.0%20CSV.zip
12. US EPA Decentralized Wastewater Program}. Septic Systems Fact Sheet. 2008 Oct.
13. Hill DT, Larsen DA. Using geographic information systems to link population estimates to wastewater surveillance data in New York State, USA. PLOS Glob Public Health. 2023;3: e0001062.
14. U.S. Environmental Protection Agency, Office of Water. Report to Congress on The Prevalence Throughout the U.S. of Low- and Moderate-Income Households Without Access to a Treatment Works and The Use by States of Assistance under Section 603(c)(12) of the Federal Water Pollution Control Act. U.S. Environmental Protection Agency; 2021 Jul. Available: https://www.epa.gov/system/files/documents/2022-01/low-mod-income-without-treatment_report-to-congress.pdf
15. U.S. Census Bureau. AHS 2019 National Public Use File. Available: https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20National%20PUF%20v1.1%20CSV.zip

16. U.S. Census Bureau. AHS 2019 Metropolitan Public Use File. Available: https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20Metropolitan%20PUF%20v1.0%20CSV.zip
17. U.S. Census Bureau, Department of Commerce Department of Housing and Urban Development. 2019 AHS Integrated National Sample: Sample Design, Weighting, and Error Estimation. U.S. Census Bureau, Department of Commerce Department of Housing and Urban Development; 2020 Sep. Available: https://www2.census.gov/programs-surveys/ahs/2019/2019%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf
18. U.S. Census Bureau, Department of Commerce Department of Housing and Urban Development. 2019 AHS Metropolitan Sample: Sample Design, Weighting, and Error Estimation. U.S. Census Bureau, Department of Commerce Department of Housing and Urban Development; 2020 Sep. Available: https://www2.census.gov/programs-surveys/ahs/2019/2019%20AHS%20Metropolitan%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf
19. U.S. Census Bureau, Department of Commerce Department of Housing and Urban Development. 2021 AHS Integrated National Sample: Sample Design, Weighting, and Error Estimation. U.S. Census Bureau, Department of Commerce Department of Housing and Urban Development; 2022 Aug. Available: https://www2.census.gov/programs-surveys/ahs/2021/2021%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf
20. US Census Bureau. American Community Survey Information Guide. [cited 5 Jan 2023]. Available: https://www.census.gov/programs-surveys/acs/library/information-guide.html
21. General Facts and Statistics about Wastewater in Florida. [cited 10 Feb 2023]. Available: https://floridadep.gov/water/domestic-wastewater/content/general-facts-and-statistics-about-wastewater-florida

22. Florida Department of Health. Onsite Sewage Locations in Florida - 2012. Available: https://fgdl.org/zips/metadata/xml/septic_jun12.xml
23. Environmental Protection Agency. Clean Watersheds Needs Survey Detailed Scope and Methods. Environmental Protection Agency; 2012. Available: https://www.epa.gov/sites/default/files/2015-10/documents/cwns_2012_detailed_scope_and_methods-508.pdf
24. Agency USEP. Clean watersheds needs survey 2012 report to congress. US Environmental Protection Agency; 2016.
25. U.S. Census Bureau. Guidance on Using 2020 Island Areas Censuses Data. 20 Oct 2022 [cited 31 Mar 2023]. Available: https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/2020-iac-guidance.pdf
26. U.S. Census Bureau. 2020 Island Areas Censuses (IAC) Demographic Profile Summary File. U.S. Census Bureau; 2022 Oct. Available: https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/demographic_profile/2020-iac-dpsf-technical-documentation.pdf
27. U.S. Department of Housing and Urban Development and U.S. Census Bureau. 2013 AHS Definitions. In: 2013 AHS Definitions [Internet]. [cited 9 Feb 2023]. Available: https://www2.census.gov/programs-surveys/ahs/2013/2013%20AHS%20Definitions.pdf
28. TIGER/Line® Shapefiles. [cited 25 Apr 2023]. Available: https://www.census.gov/cgi-bin/geo/shapefiles/index.php
29. U.S. Census Bureau. Metropolitan and Micropolitan About. In: Metropolitan and Micropolitan About [Internet]. 22 Nov 2021 [cited 9 Feb 2023]. Available: https://www.census.gov/programs-surveys/metro-micro/about.html
30. Schwartz MB. 2007 American Community Survey: A Comparison to Selected Housing and Financial Characteristics for the United States from the 2007 American Housing Survey.

Housing and Household Economic Statistics Division U.S. Census Bureau; Available: https://www.huduser.gov/portal/datasets/ahs/acs-ahscomparisonreport.pdf

31. U.S. Department of Housing and Urban Development and U.S. Census Bureau. 2021 AHS Definitions. In: 2021 AHS Definitions [Internet]. [cited 22 May 2023]. Available: https://www2.census.gov/programs-surveys/ahs/2021/2021%20AHS%20Definitions.pdf
32. U.S. Census Bureau. AHS 2013 National Public Use File. Available: https://www2.census.gov/programs-surveys/ahs/2013/AHS%202013%20National%20PUF%20v2.0%20CSV.zip
33. Indian Health Service; Office of Environmental Health and Engineering; Division of Sanitation Facilities Construction. Annual Report To the Congress of the United States On Sanitation Deficiency Levels for Indian Homes and Communities: Fiscal Year 2019. Rockville (MD): Indian Health Service (US); 2020.
34. Environmental Protection Agency. EPA's Proposal: Justification for Requested Revision, Addition or Deletion to Questions on the American Community Survey (ACS). 2018. Available: https://foiaonline.gov/foiaonline/action/public/submissionDetails?trackingNumber=EPA-HQ-2018-010667&type=request
35. Duvallet C, Wu F, McElroy KA, Imakaev M, Endo N, Xiao A, et al. Nationwide Trends in COVID-19 Cases and SARS-CoV-2 RNA Wastewater Concentrations in the United States. ACS EST Water. 2022. doi:10.1021/acsestwater.1c00434
36. Medina CY, Kadonsky KF, Roman FA Jr, Tariqi AQ, Sinclair RG, D'Aoust PM, et al. The need of an environmental justice approach for wastewater based epidemiology for rural and disadvantaged communities: A review in California. Curr Opin Environ Sci Health. 2022;27: 100348.

651 37. Haak L, Delic B, Li L, Guarin T, Mazurowski L, Dastjerdi NG, et al. Spatial and temporal
652 variability and data bias in wastewater surveillance of SARS-CoV-2 in a sewer system. Sci
653 Total Environ. 2022;805: 150390.
654 38. Adhikari S, Halden RU. Opportunities and limits of wastewater-based epidemiology for
655 tracking global health and attainment of UN sustainable development goals. Environ Int.
656 2022;163: 107217.
657 39. Jakariya M, Ahmed F, Islam MA, Al Marzan A, Hasan MN, Hossain M, et al. Wastewater-
658 based epidemiological surveillance to monitor the prevalence of SARS-CoV-2 in developing
659 countries with onsite sanitation facilities. Environ Pollut. 2022;311: 119679.
660 40. Gerrity D, Papp K, Dickenson E, Ejjada M, Marti E, Quinones O, et al. Characterizing the
661 chemical and microbial fingerprint of unsheltered homelessness in an urban watershed. Sci
662 Total Environ. 2022;840: 156714.
663 41. Tokajian S, Moghnieh R, Salloum T, Arabaghian H, Alousi S, Moussa J, et al. Extended-
664 spectrum β-lactamase-producing Escherichia coli in wastewaters and refugee camp in
665 Lebanon. Future Microbiol. 2018;13: 81–95.
666 42. Gwenzi W. Wastewater, waste, and water-based epidemiology (WWW-BE): A novel
667 hypothesis and decision-support tool to unravel COVID-19 in low-income settings? Sci
668 Total Environ. 2022;806: 150680.
669 43. U.S. Department of Housing and Urban Development and U.S. Census Bureau. Appendix
670 B: Sample Design and Weighting: 2013. U.S. Department of Housing and Urban
671 Development and U.S. Census Bureau; 2013. Available:
672 https://www2.census.gov/programs-
673 surveys/ahs/2013/2013%20AHS%20National%20Sample%20Design%20and%20Weightin
674 g.pdf
675 44. Onsite Sewage FAQ - Permitting. [cited 10 Feb 2023]. Available:
676 https://floridadep.gov/water/onsite-sewage/content/onsite-sewage-faq-permitting

45. Florida Department of Health. Septic System Owner's Guide. In: floridahealth.gov [Internet]. [cited 10 Feb 2023]. Available: https://www.floridahealth.gov/environmental-health/onsite-sewage/_documents/septic-folder-8x11-link2.pdf
46. Utah Department of Environmental Quality Water Quality. Municipal Wastewater Planning Program (MWPP). In: Municipal Wastewater Planning Program (MWPP) [Internet]. 3 Feb 2023 [cited 10 Feb 2023]. Available: https://deq.utah.gov/water-quality/municipal-wastewater-planning-program-mwpp
47. Campbell H. 2021 MWPP Survey. In: 2021 MWPP Survey [Internet]. 27 Oct 2021 [cited 10 Feb 2023]. Available: https://documents.deq.utah.gov/water-quality/engineering/municipal-wastewater-planning-program/DWQ-2021-026722.pdf
48. Minnesota Pollution Control Agency. 2021 Wastewater Infrastructure Needs Instructions and Frequently Asked Questions (FAQ). Apr 2021 [cited 31 Mar 2023]. Available: https://www.pca.state.mn.us/sites/default/files/wq-wwtp3-06.pdf
49. Robinson C. 2017 SSTS Annual Report Subsurface Sewage Treatment Systems in Minnesota. Minnesota Pollution Control Agency; 2018 Apr. Available: https://www.pca.state.mn.us/sites/default/files/wq-wwists1-58.pdf
50. U.S. Environmental Protection Agency. 2012 Clean Watersheds Needs Survey. Available: https://ordspub.epa.gov/ords/cwns2012/f?p=cwns2012:25
51. Weidhaas J, Aanderud ZT, Roper DK, VanDerslice J, Gaddis EB, Ostermiller J, et al. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. Sci Total Environ. 2021;775: 145790.
52. U.s E. WaterSense Statistics and Facts. 11 May 2022 [cited 3 Apr 2023]. Available: https://www.epa.gov/watersense/statistics-and-facts

Supporting information

S1 Appendix. Supplementary method

S2 Appendix. Supplementary analyses. Compilation of data, assessment of data completeness, sewer connectivity in U.S. Island Areas, supplementary figures, and supplementary tables.

S1 Appendix. Supplementary methods.

Dataset descriptions and data cleaning

U.S. Census Bureau American Housing Survey 2019 and 2021

The American Housing Survey (AHS) National and Metropolitan Public Use Files were downloaded from Ref. [11,15,16]. The sample design, weighting, and error estimation survey is described in Refs. [17–19]. Briefly, the AHS is a biannual survey that assesses housing characteristics. Two samples of households are chosen every survey cycle, the Integrated National Sample and the Independent Metropolitan Sample. The Integrated National Sample consists of nationally representative households and includes an oversample of each of 15 largest metropolitan areas. The Independent Metropolitan Sample consists of representative households in 20 large metropolitan areas that are representative of the next largest 50 metropolitan areas. Ten of the 20 large metropolitan areas are sampled every other survey cycle (10 metropolitan areas surveyed every 2 years). The Integrated National Sample in 2019 included 86,257 selected representative housing units, of which about 63,500 met the AHS definition of a housing unit and were able to be interviewed, and in 2021 included 95,295 selected representative housing units, of which about 64,100 were interviewed. The Integrated Metropolitan Samples in 2019 and 2021 included around 3000 selected representative housing units in each metropolitan area, of which around 1000-3000 were eligible and were able to be interviewed (same criteria as in the Integrated National Sample).

Geographic data of each household available in the Public Use File included the census division; core based statistical area (CBSA) for the 15 largest metropolitan areas and the 20 large metropolitan areas in the metropolitan data. All other households are classified as in “all other metropolitan areas” or “not in a metropolitan area” (all other locations).

The race, ethnicity, and age of the householder (the person or one of the persons who is an owner or renter of the unit) was used to categorize the race, ethnicity, and age of the household. Due to the low numbers of households where the householder responded as having more than one race, we focused on households where householders identified as having a single race.

U.S. Census Bureau American Housing Survey National Sample 2013

The American Housing Survey National (2013) Use File was downloaded from Ref. [32]. The sample design, weighting, and error estimation survey is described in Ref [43]. The National Sample included 84,400 selected housing units representative nationally and including a supplemental sample of 15,553 housing units in the Chicago, Detroit, New York City, Northern New Jersey, and Philadelphia metropolitan statistical areas, of which about 71,600 met the definition of a housing unit and were able to be interviewed.

Geographic data of each household available in the Public Use File included the census division; census region; central city, urban, or rural status; and select metropolitan area codes. All other housing units are classified into “not in a metropolitan area” or had their code suppressed for confidentiality [27].

Florida Department of Health onsite sewage treatment and disposal system inspections

The locations of onsite sewage treatment and disposal systems inspected by Florida Department of Health reported June 2012 was downloaded from Ref. [22]. As of 2022, an inspection permit is required for construction [44] and inspection is recommended every 3 to 5 years after [45]. The number of systems inspected that are active at the latest inspection (after removing repeat inspections) is 601,430. The years of inspection are primarily 1998-2011, with some outliers as early as 1900, which may be old septic systems or typos, and as late as 2200, which are likely typos.

Utah Municipal Wastewater Planning Program (MWPP) Survey 2021

Data from the Utah Municipal Wastewater Planning Survey was provided by the Utah Department of Environment Quality, Division of Water Quality. The dataset includes an estimate of the population receiving collection by each utility that owns or operates a sanitary sewerage system and its latitude and longitude. The data gathered from treatment plant operators and participation is reported as mandatory [46], although only 71% of contacts responded [47]. We filtered the data to include any facility that has collection (this included collection only, treatment and collection, or small lagoons which are often collection and treatment). Manual correction of facility longitude and latitude is described in the supplementary methods.

Manual editing of facility longitude and latitude was done for 5 facilities' that did not fall within the state of Utah when combining with the 2021 TIGER shapefile (BRIAN HEAD TOWN, CANYONG LAND IMPROVEMENT DISTRICT, HILL AFB (AMERICAN WATER), JORDANELLE SSD, PLAIN CITY) and 1 facility that did not report a latitude and longitude (PANGUITCH LAKE S. S. D.). For these facilities, the latitude and longitude were manually

edited to be that of the location found in a Google Maps search of the facility name. If the Google Maps search found a city, then an arbitrarily chosen location in that city was used to obtain a latitude and longitude.

Minnesota Wastewater Infrastructure Needs Survey (WINS) 2021

Data for the Minnesota Wastewater Infrastructure Needs Survey [48] was provided by the Minnesota Pollution Control Agency. Communities (places and townships) responded to the question “Does your community have a collection system?” with “yes”, “no”, or no response. The survey did not receive a response from every community in Minnesota.

The Census Designated Place (CDP) name was obtained from the reported community name with manual changes reported in **Table S *2***.

Minnesota Subsurface Sewage Treatment Systems 2017

The Minnesota Subsurface Sewage Treatment Systems (SSTS) dataset reported the number of SSTS by local government units with a known SSTS program. Data was taken from Ref. [49]. Two hundred and eleven out of 2018 known SSTS programs submitted an annual report. Not all SSTS programs were able to be identified and contacts were not always provided. Fourteen reported having zero SSTS within their jurisdiction despite permitting SSTS in 2017. Data were reported to sometimes be estimates.

U.S. Environmental Protection Agency Clean Watersheds Needs Survey 2012

Data for the most recent Environmental Protection Agency (EPA) Clean Watersheds Needs Survey (CWNS) survey (2012) [24] was downloaded from [50] and extracted from a Microsoft

Access database file to csv files using the mdbtools package (https://github.com/mdbtools/mdbtools). The survey is administered by the EPA and the States to assess funding needs for unfunded capital costs of treatment works projects that (1) address a water quality or related public health problem existing as of January 1, 2012, or is expected to occur within the next 20 years and (2) meet the CWNS documentation criteria. CWNS documentation criteria includes (1) description and location of the problem, (2) site-specific solution to the problem, and (3) detailed cost information for implementing the solution. Data is collected about (1) publicly owned wastewater collection and treatment facilities, (2) combined sewer overflow control facilities, (3) stormwater management activities, and (4) decentralized wastewater treatment facilities. Data is collected on (a) estimated needs, costs and technical information, (b) facility location and contact information, and (c) facility population served, flow, effluent, and unit process information. Supporting documentation was required for the entered data with some exceptions (for instance small communities could use a more simple form), and quality control was performed by reviewers to compare documentation with data entered in the system to ensure consistency of technical and cost data [23]. Data entry occurred from January to December 2012.

There was large variability in the level of effort and resources that each state put into the survey. New York, California, Florida, New Jersey, Maryland, Iowa, Minnesota, and Michigan likely invested heavily in participating in the survey or already had similar state-level data collection systems in place for getting comprehensive responses [22,48,49].

Facilities not included in the dataset are those in South Carolina, the Northern Mariana Islands, and American Samoa as they did not participate in the survey; facilities whose projects did not have documented solutions or cost estimates; privately owned wastewater facilities that serve privately owned industrial facilities, military installations, national parks, or other federal facilities;

facilities on tribal lands and in Alaskan Native Villages, which were separately surveyed by the Indian Health Service; and facilities with projects that had existing funding prior to January 1, 2012. Additionally, larger facilities were more likely to be captured by the survey due to the resource requirements for obtaining the right documentation. To address this bias, the survey had decreased documentation needs for small communities (population <=10,000); despite this, the small community facilities were still underrepresented (pg. 2 of Ref. [24]).

We filtered the facilities to the current facilities of type “collection: combined sewers” or “collection: separate sewers”. The population receiving collection is taken from the reported total number of residents who are connected to a sewer system which empties into a treatment plant. This does not include non-resident populations, populations served by acceptable decentralized wastewater treatment systems, or populations connected to sewers that do not discharge to a treatment plant.

U.S. Census Bureau Island Areas Decennial Survey 2020

The Island Areas Decennial Survey surveys housing, social, and economic information for all housing units in Guam, the Northern Mariana Islands, the Virgin Islands, and American Samoa [26]. Data of general demographic characteristics (DP01), selected economic characteristics (DP03), and selected housing characteristics (DP04) were downloaded from https://data.census.gov/. Housing characteristics included data for both occupied and vacant housing units.

U.S. Census Bureau American Community Survey

The American Community Survey (ACS) is a monthly survey of a representative subsample of addresses in the US [20] on characteristics of populations and households. The ACS 5 year

estimates (average over 5 years, less subject to fluctuations) for selected social characteristics (DP02), selected economic characteristics (DP03), and demographic and housing estimates (DP05) stratified county or county subdivision for select years were downloaded from https://data.census.gov/.

Data analysis and visualization

Florida Department of Health onsite sewage treatment and disposal system inspections dataset analysis

The number of septic tanks in each county subdivision was calculated as the number of septic permits whose latitude and longitude occurred inside the boundaries of each county subdivision according to the 2012 U.S. Census Bureau county subdivision TIGER shapefile. To calculate the fraction of a county subdivision connected to septic tanks, the number of septic tanks in each county subdivision was divided by the number of households in the county subdivision as reported by the 2012 American Community Survey 5-year estimate. Only county subdivisions with at least 5 households and 20 population size were used.

Utah Municipal Wastewater Planning Program Survey analysis

The population in each county receiving collection was calculated by summing the population receiving collection across all reported facilities in that county. The population in each county subdivision receiving collection was calculated by summing the population receiving collection across all reported facilities whose latitude and longitude fell within that county subdivision's boundaries from the 2021 TIGER shapefile. To calculate the fraction of a county or county subdivision receiving sewer collection, the population receiving collection was divided by the population size of each county or county subdivision reported by the 2021 American Community

Survey 5-year estimate. Only county subdivisions with at least 5 households and 20 population size were used.

U.S. Environmental Protection Agency Clean Watersheds Needs Survey analysis

To compare the EPA CWNS and AHS data in Core Based Statistical Areas (CBSA), counties were categorized into CBSAs using the February 2013 U.S. Census Bureau file. To calculate the fraction of a county receiving collection, the population size receiving collection in that county was divided by the population size of each county reported by the 2012 American Community Survey 5-year estimate. We flagged counties that had no facilities reported, which could either be because there were no facilities in that county or because of a lack of reporting, and counties that had >100% sewer connectivity by population (potentially due to misreporting or collection occurring across multiple counties) were set to 100% connectivity.

U.S. Census Bureau Island Areas Decennial Survey analysis

To calculate the fraction of a Census Designated Place (CDP) connected to sewers, the number of households connected to sewers was divided by the total number of households in a CDP. Only CDPs with at least 5 households and 20 population size were used.

Generation of maps

TIGER shapefiles from the U.S. Census Bureau in the closest matching year to the dataset were used for generating maps. Maps were generated using the geopandas package (version 0.12.2) in Python (version 3.11.0) and were displayed in the Mercator projection.

Simulations of interacting populations and wastewater sampling

Two interacting populations are modeled in a compartmental model with susceptible, infected, and recovered (SIR) compartments for each population. The number of susceptible, infected, and recovered individuals in population A (S_A, I_A, and R_A, respectively) and in population B (S_B, I_B, and R_B, respectively) are described by:

$$\frac{dS_A}{dt} = -\beta_A(1-\varepsilon)I_A\frac{S_A}{N_A} - \beta_B\varepsilon I_B\frac{S_A}{N_A}$$

$$\frac{dS_B}{dt} = -\beta_B(1-\varepsilon)I_B\frac{S_B}{N_B} - \beta_A\varepsilon I_A\frac{S_B}{N_B}$$

$$\frac{dI_A}{dt} = \beta_{AA}(1-\varepsilon)I_A\frac{S_A}{N_A} + \beta_{BA}\varepsilon I_B\frac{S_A}{N_A} - \gamma_I I_A$$

$$\frac{dI_B}{dt} = \beta_{BB}(1-\varepsilon)I_B\frac{S_B}{N_B} + \beta_{AB}\varepsilon I_A\frac{S_B}{N_B} - \gamma_I I_B$$

$$\frac{dR_A}{dt} = \gamma_I I_A$$

$$\frac{dR_B}{dt} = \gamma_I I_B$$

where N_A and N_B are the number of individuals in populations A and B respectively, β_A and β_B are the overall contact rates of populations A and B respectively times the probability of infection given contact, γ_I is the rate of recovery. ε is the fraction of the total contacts made by individuals in a population that occurs with individuals in the other population and it describes the interaction strength between the two populations. ε can take on values between 0 and 1, where ε=0 indicates that all interactions occur within each population and none between populations, and ε=1 indicates that all interactions occur between the populations and none within

populations. The base parameters are set at N_A=N_B=5000, γ_I=0.18 inverse days, $R_0^A = R_0^B = \frac{\beta_A}{\gamma_I} = \frac{\beta_B}{\gamma_I} = 1.5$, which represent an outbreak similar to COVID-19, but the parameters are varied in sensitivity analyses. The initial condition was a single infected individual in population A (I_A=1) and all other individuals were susceptible (S_A=N-1, S_B=N, I_B=0, R_A=0, R_B=0). The differential equations are solved deterministically using the integrate.odeint function in the scipy package (version 1.6.2) in Python using a timestep of 0.1 days.

The concentration C of the pathogen in wastewater is given by

$$C = \frac{n_A + n_B}{v_A + v_B}$$

where n_A and n_B are the number of copies of pathogen genetic material (DNA or RNA) shed per day by population A and B, respectively, into the sampled wastewater and v_A and v_B are the volume of wastewater produced per day by population A and B, respectively, into the sampled wastewater.

The number of copies of pathogen genetic material, n_A and n_B, shed per day by population A and B, respectively, into the sampled wastewater is given by

$$n_A = nI_A f_A$$

$$n_B = nI_B f_B$$

Where n is the number of copies of pathogen genetic material shed per infected person per day (assumed to be the same for all infected individuals) and f_A and f_B are the fraction of total shed

pathogen genetic material by population A and B respectively that is sampled by wastewater (determined by the sewer connectivity of each of the populations).

The volume of wastewater, v_A and v_B, produced per day by population A and B, respectively, into the sampled wastewater is given by

$$v_A = vN_A f_A$$

$$v_B = vN_B f_B$$

where v is the volume of wastewater produced per person per day (assumed to be the same for all individuals). The value used for n was $5 \times 10^{6.7}$ gene copies/person/day (estimate taken from Ref. [51] by multiplying $10^{4.7}$ gene copies/mL feces by 500 mL feces/day/person) and for v was 20 liters wastewater/person/day (per person daily wastewater usage approximated from per household daily wastewater usage from Ref. [52] assuming 4 people per household).

In addition to the assumptions mentioned above, the model makes the following additional assumptions: The probability of an infected individual shedding pathogen genetic material to wastewater is 1. Given that there is pathogen genetic material present in wastewater, the probability of detecting it is 1. Shedding of pathogen genetic material occurs only when an individual is infected. There are no stormwater contributions to wastewater to dilute the pathogen concentration.

To calculate the time of the maximum number of infections in a population and the time of the maximum concentration of pathogen gene content in wastewater data, we numerically calculated the time derivative at the midpoints of the time points as the difference between consecutive datapoints divided by the time interval and took the first time point at which the derivative crossed 0.

969

970

S2 Appendix. Supplementary analyses.

Compilation of data and assessment of data completeness

As the AHS does not provide finer scale geographic information than the census division for households outside of the 35 selected large metropolitan areas, we used the American Community Survey, which reports the percentage of occupied housing units lacking complete plumbing facilities by county and which allowed us to define the lower bound on the percentage of the population not connected to sewers. Additionally, we identified several state-level datasets, including locations of permitted septic tanks in Florida, sewer locations and populations served in Utah and Minnesota, and the 2012 EPA Clean Watersheds Needs Survey of sewer locations and populations served for 49 states (South Carolina did not participate). However, many of these state-level datasets are incomplete or potentially biased and thus only allow qualitative analyses. Finally, the 2020 US Census' Island Areas Decennial survey provided information the percentage of households connected to sewer by Census Designated Place in the island areas of American Samoa, Guam, the Northern Mariana Islands, and the Virgin Islands (see **Error! Reference source not found.** and Table S 3 for a summary of datasets).

We checked the consistency of the Florida Department of Health onsite sewage treatment and disposal systems inspections data reported in 2012 with other sources. We noted that an estimated 7% of households across the state were connected to septic tanks, but that this was substantially less than the Florida Department of Environmental Protection's estimate that approximately one third of Florida's population uses septic tanks [21] even after accounting for potential correlations of septic usage with household size (approximate increase in household size of 10% for a 50% increase in septic connectivity (**Fig S 5**) cannot explain the almost 5 time

difference in septic connectivity across the two sources), suggesting that more households may be on septic tanks, and thus less on sewer systems. This discrepancy may exist because not every septic tank had a permit recorded in the state system, because there were changes since 2012, or because of uncertainty in the Florida Department of Environmental Protection's estimate. Thus, we focused our analysis on relative comparisons across county subdivisions (Fig 4).

The 2012 EPA Clean Watersheds Needs Survey dataset includes voluntary submissions of the locations of publicly owned wastewater collection and treatment facilities and their estimated population served. The following states had among the highest reported needs and survey participation level (pg. 6 and Table A-1 of Ref. [24]) and were the focus of our analyses: New York, California, Florida, New Jersey, Maryland, Iowa, Minnesota, and Michigan. Comparison of Florida in the EPA and Florida Department of Public Health datasets showed consistent spatial trends (less connection to sewers and more connection to septic systems in the panhandle and inland) (Fig 4 and Fig S ***9***). Additionally, in both datasets, counties with more connection to sewers are more Asian (EPA: Pearson $r = 0.53$ ($p = 4.2\times10^{-4}$), Florida Department of Public Health: Pearson $r = 0.31$ ($p = 3.2\times10^{-7}$)). Comparison of Minnesota in the EPA dataset, Minnesota Wastewater Infrastructure Needs Survey, and Minnesota SSTS datasets also showed consistent spatial trends (more connection to sewers and less connection to subsurface sewage treatment systems around Minneapolis) (**Fig S 11**). We found a significant correlation between the AHS and EPA datasets for the largest 35 metropolitan areas in the states with good data collection (**Fig S 12**), despite deviations from a one-to-one correspondence suggesting overall systematic biases in the data.

The 2021 Utah Municipal Wastewater Planning Survey estimated the populations receiving collection by municipal utilities that owned or operated a sanitary sewerage system. The

purpose of the survey is to aid in overseeing and communicating with the wastewater industry in Utah. We noticed that these data may have not be quantitatively accurate but are likely qualitatively informative. First, large swaths of the state appeared not to receive wastewater collection (**Fig S 13**), but this may partly reflect that only 71% of contacts responding to the survey [47]. Second, in some county subdivisions, >100% of the population appeared connected to sewers, likely due to errors in estimation, differences in how combined treatment and collection facilities reported population estimates, or collection populations resided in neighboring county subdivisions. While the 5 county subdivisions with the highest percentages of American Indian and Alaska Natives (~20-100%) were reported to not be connected to sewers at all in this dataset, we believe this is because Indian reservations are not included in the dataset (as they are under federal, not state jurisdiction). In fact, all 5 of these counties (Uintah and Ouray, Casa del Eco Mesa-White Mesa, West Juab, Oljato, Blanding) include Indian reservations. Additionally, the county subdivision with the highest percentage of Black or African American; Native Hawaiian and Other Pacific Islander; or Hispanic are in county subdivisions has nobody receiving collection (**Fig S 14**); however, we cannot rule out that this is due to underreporting from these county subdivisions.

The 2021 Minnesota Wastewater Infrastructure Needs Survey reported whether communities have a collection system. The data from this survey was geographically sparse, as communities (the level of aggregation of the dataset) often directly mapped to a census designated place (CDP), and CDPs do not cover the entire state (**Fig S 11**a). Consequently, the sewer connectivity of the rural areas between census designated places could not be determined from this dataset. Additionally, not all communities responded to the survey. Of the 523 communities that participated in the survey, all but 11 had a collection system. The geographic distribution of sewer connectivity qualitatively matched that of the 2017 Subsurface Sewage Treatment Systems (SSTS) Annual Report of the number of reported subsurface sewage treatment

systems (septic tanks) by county in Minnesota (**Fig S 11**b); however, 7 out of the 218 contacted local government units did not respond, 14 reported having zero SSTS within their jurisdiction despite permitting SSTS and not all SSTS programs were able to be identified to be contacted.

Given the data incompleteness and biases for the Utah and Minnesota datasets, we excluded them from the analyses in the main text.

Sewer connectivity in the U.S. Island Areas

From the 2020 US Census Island Areas Decennial Survey of household characteristics, the overall levels of household sewer connection in the island areas were lower than in the states (**Fig S 15**). In Guam, the Northern Mariana Islands, the Virgin Islands, and American Samoa, the overall connection to sewer systems across each island was 65%, 51%, 66%, and 52% of households, respectively (compared to an overall 83% connection across mainland US). To ask whether there was strong spatial variability across each island, we looked at the median household connection by Census Designated Place (CDP). A median of 64%, 32%, 70%, and 13% households by CDP in Guam, the Northern Mariana Islands, the Virgin Islands, and American Samoa, respectively, were connected to sewers. This suggests that sewer connectivity is particularly spatially concentrated in the Northern Mariana Islands and American Samoa (difference in median household connection by CDP and overall island connection), although all island areas showed some spatial variability in connectivity when visually assessing maps (**Fig S 16**). The only significant correlation of sewer connectivity with demographic or economic characteristics was in American Samoa, where CDPs with a higher percentage of Black or African Americans were more connected to sewers (Pearson $r = 0.38$, q-value = 0.01).

Supplementary figures

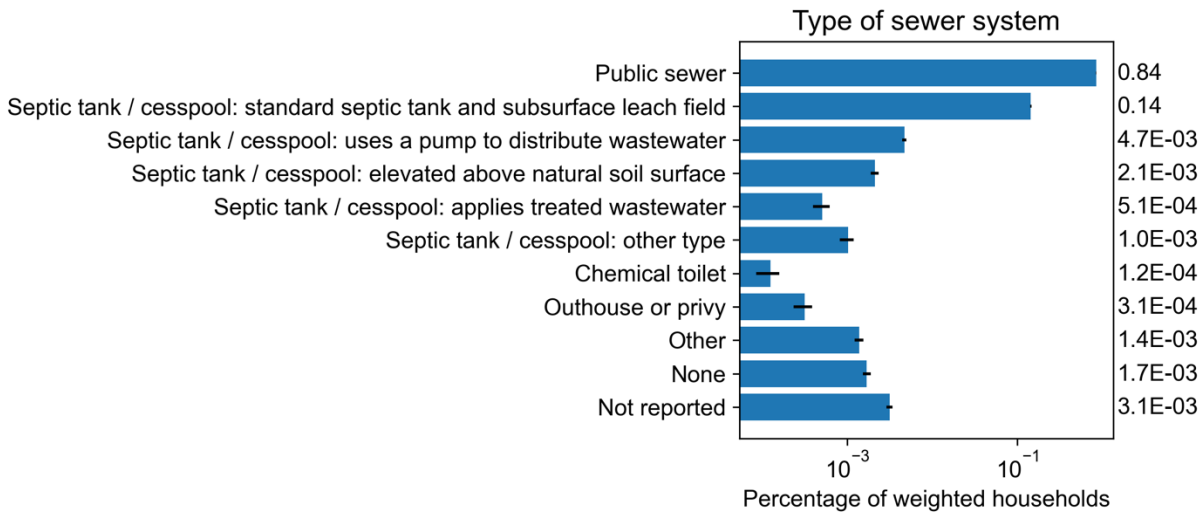
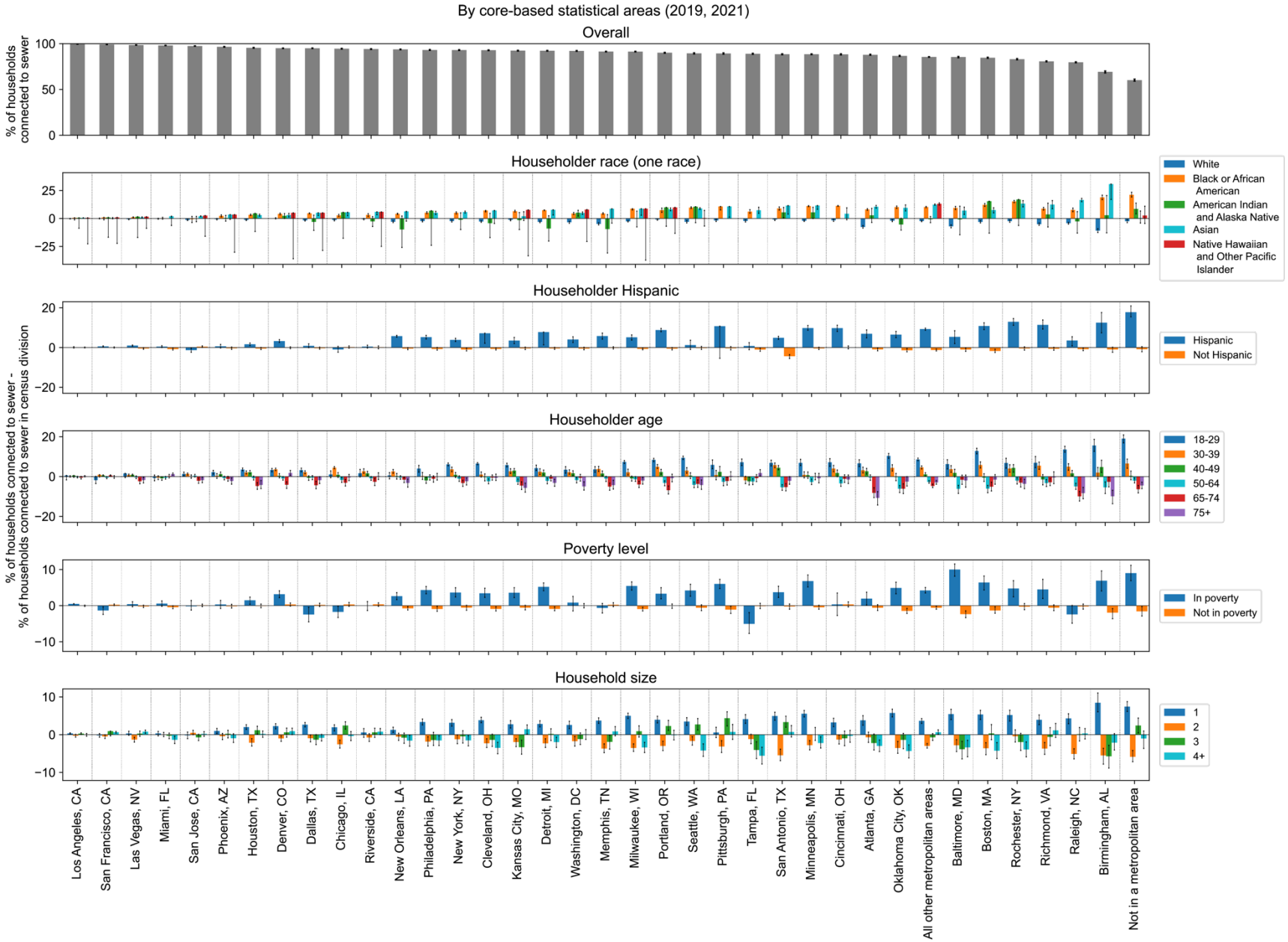


Fig S 1: **Percentage of households connected to various wastewater processing systems.**

Data are from 2021 American Housing Survey.


By core-based statistical areas (2019, 2021)
Overall
% of households connected to sewer
Householder race (one race)
White
Black or African American
American Indian and Alaska Native
Asian
Native Hawaiian and Other Pacific Islander
Householder Hispanic
Hispanic
Not Hispanic
Householder age
18-29
30-39
40-49
50-64
65-74
75+
Poverty level
In poverty
Not in poverty
Household size
1
2
3
4+
% of households connected to sewer - % of households connected to sewer in census division
Los Angeles, CA
San Francisco, CA
Las Vegas, NV
Miami, FL
San Jose, CA
Phoenix, AZ
Houston, TX
Denver, CO
Dallas, TX
Chicago, IL
Riverside, CA
New Orleans, LA
Philadelphia, PA
New York, NY
Cleveland, OH
Kansas City, MO
Detroit, MI
Washington, DC
Memphis, TN
Milwaukee, WI
Portland, OR
Seattle, WA
Pittsburgh, PA
Tampa, FL
San Antonio, TX
Minneapolis, MN
Cincinnati, OH
Atlanta, GA
Oklahoma City, OK
All other metropolitan areas
Baltimore, MD
Boston, MA
Rochester, NY
Richmond, VA
Raleigh, NC
Birmingham, AL
Not in a metropolitan area

1074

1075 **Fig S 2*:* Top 35 metropolitan areas' connectivity to public sewers by demographics.** Data are from the 2019 and 2021 U.S.

1076 Census American Housing Survey [11,16].

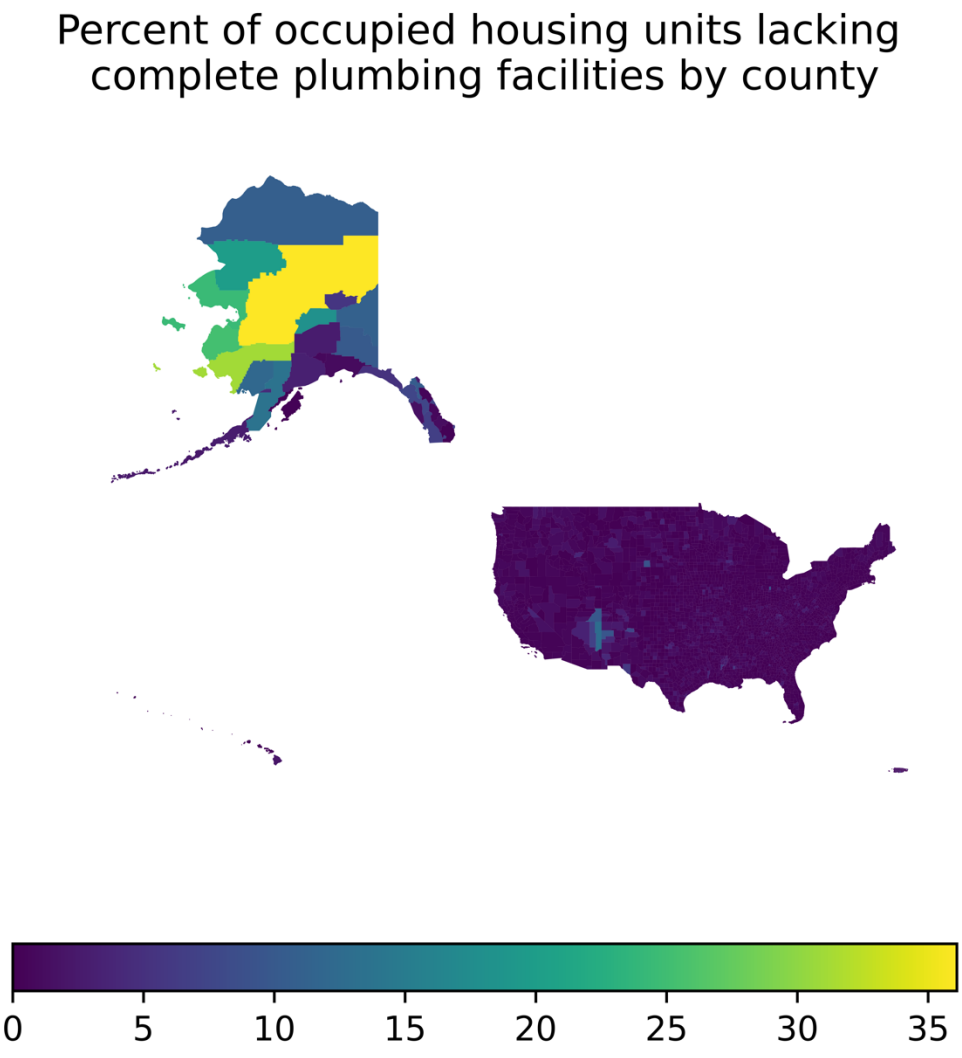


Fig S 3: Percent of occupied housing units lacking complete plumbing facilities by county. Data are from the 2021 U.S. Census American Community Survey.

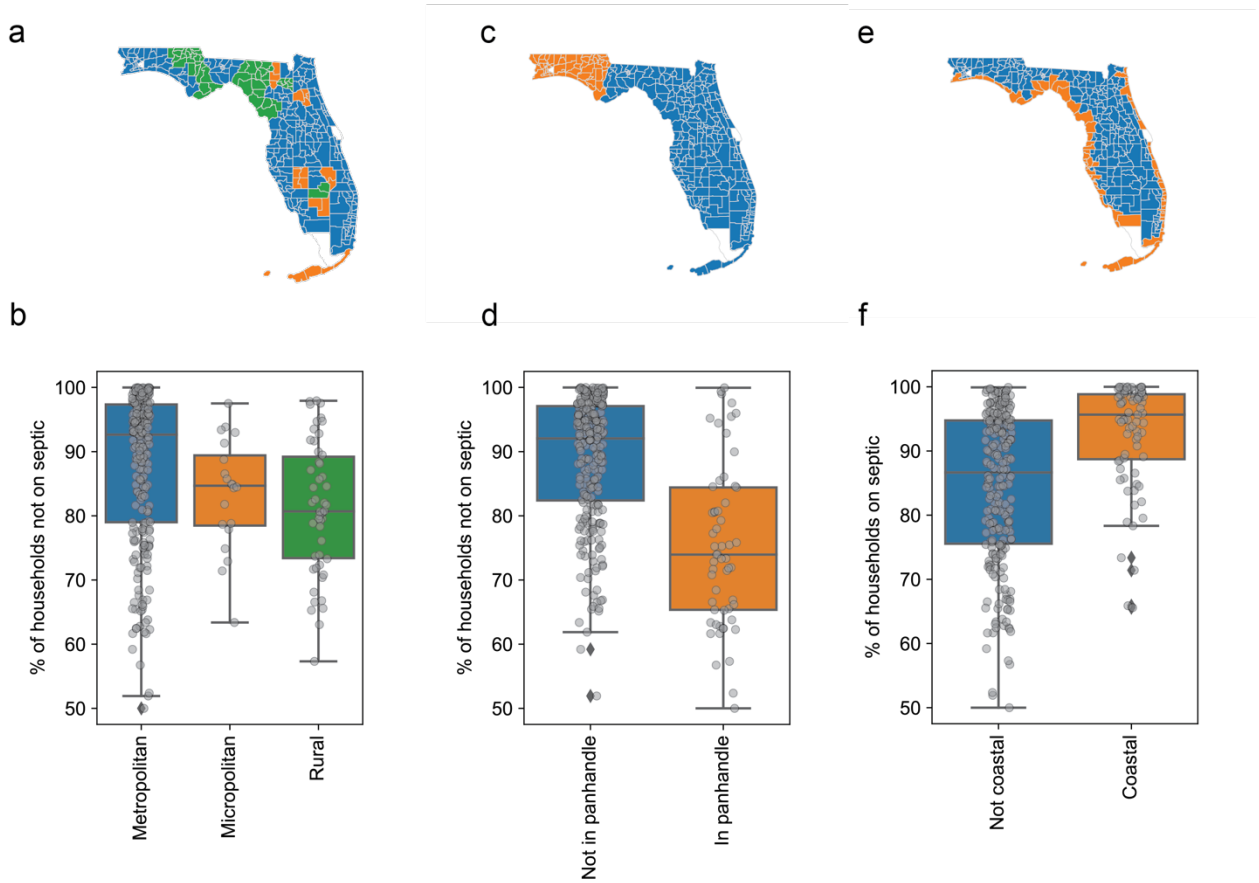


Fig S 4: Comparison of percentage of households not on septic across different geographies in Florida. Data are from the Florida Department of Health septic tank inspection permits reported in 2012. (a) Map of county subdivision colored by whether it is a Metropolitan Statistical Area (blue), Micropolitan Statistical Area (orange), or Rural Area (neither metropolitan not micropolitan, green). (b) Percentage of households not on septic in county subdivisions stratified by whether the county subdivision is a Metropolitan Statistical Area, Micropolitan Statistical Area, or Rural Area. County subdivisions that are Metropolitan Statistical Areas have a higher percentage of households not on sewer than those that are Micropolitan Statistical Areas, but the effect is not significant (p = 0.052). County subdivisions that are Micropolitan Statistical Areas have a higher percentage of households not on sewer than those that are Rural Areas, but the effect is not significant (p = 0.37). (c) Map of county subdivisions colored by whether they are always included in references to the Florida panhandle. Orange: county

subdivisions in the panhandle. Blue: county subdivisions not in the panhandle. (d) Percentage of households not on septic in county subdivisions stratified by whether the county subdivision is in the panhandle. Each point represents a county subdivision. County subdivisions in the panhandle have a significantly lower percentage of households not on septic ($p < 10^{-9}$). (e) Map of county subdivisions colored by whether they have a coastline. Orange: county subdivision with coastline. Blue: county subdivisions without coastline. (f) Percentage of households not on septic in county subdivisions stratified by coastal and not coastal county subdivisions. Each point represents a county subdivision. County subdivisions not on the coast have a significantly higher fraction of households on septic ($p<10^{-9}$).

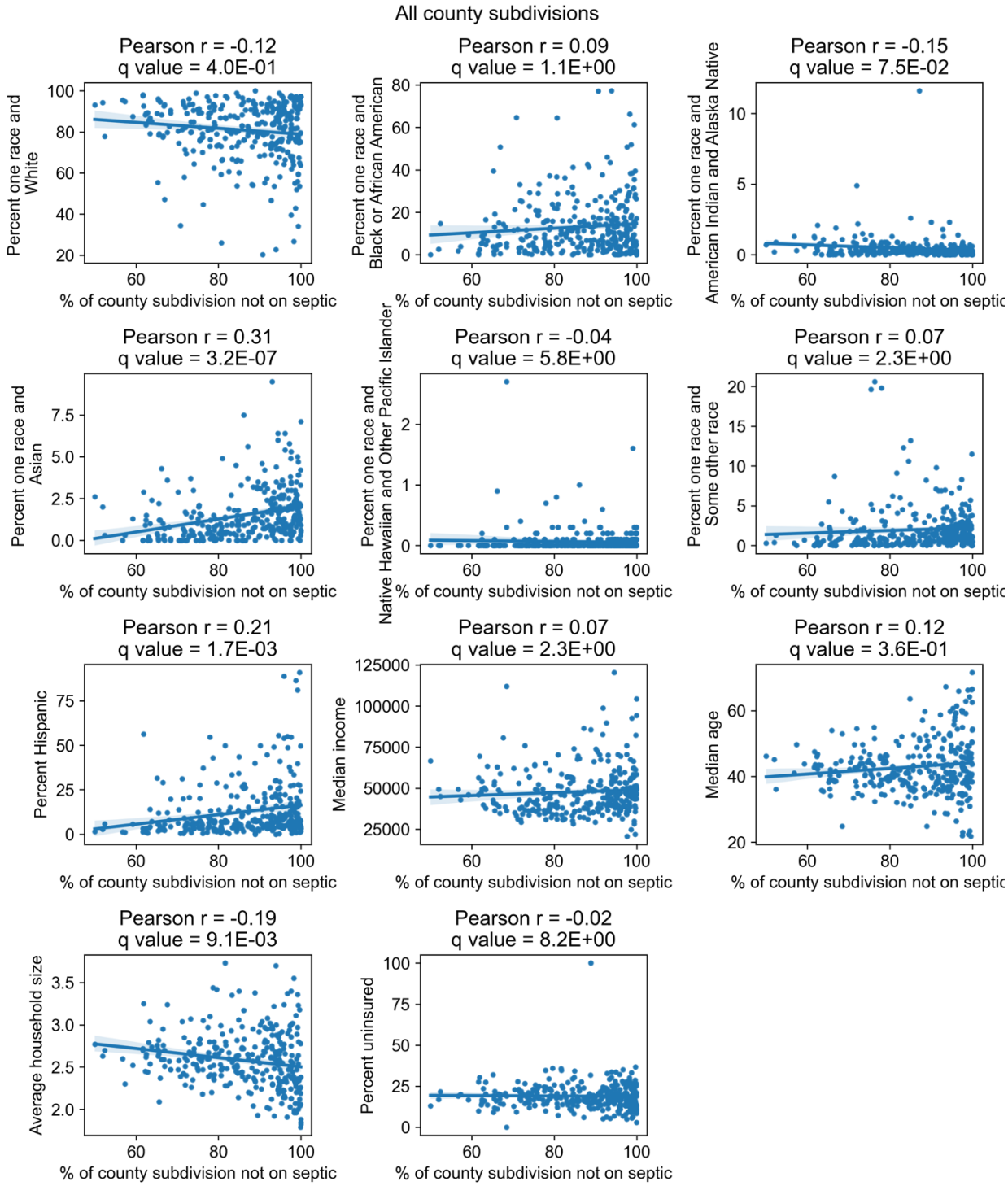

All county subdivisions
Pearson r = -0.12
q value = 4.0E-01
Percent one race and White
Pearson r = 0.09
q value = 1.1E+00
Percent one race and Black or African American
Pearson r = -0.15
q value = 7.5E-02
Percent one race and American Indian and Alaska Native
Pearson r = 0.31
q value = 3.2E-07
Percent one race and Asian
Pearson r = -0.04
q value = 5.8E+00
Percent one race and Native Hawaiian and Other Pacific Islander
Pearson r = 0.07
q value = 2.3E+00
Percent one race and Some other race
Pearson r = 0.21
q value = 1.7E-03
Percent Hispanic
Pearson r = 0.07
q value = 2.3E+00
Median income
Pearson r = 0.12
q value = 3.6E-01
Median age
Pearson r = -0.19
q value = 9.1E-03
Average household size
Pearson r = -0.02
q value = 8.2E+00
Percent uninsured
% of county subdivision not on septic

Fig S 5*:* Correlation of percentage of households in county subdivisions of Florida *not* connected to septic tanks with demographic and economic factors. Data are from the Florida Department of Health septic tank inspection permits reported in 2012. Each point represents a county subdivision. Only county subdivisions with at least 5 households and 20 population size were included.

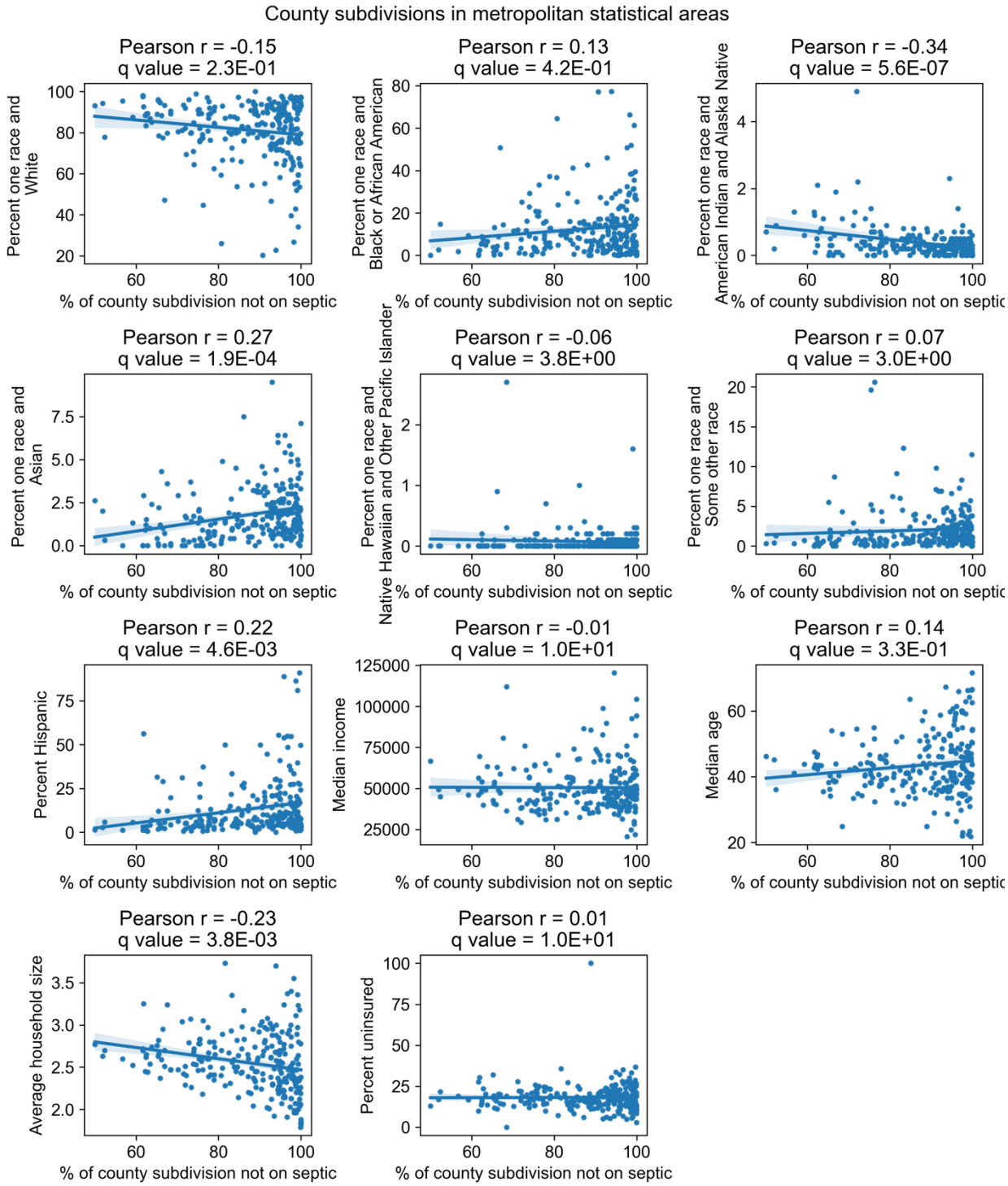


Fig S 6: Correlation of percentage of households in county subdivisions of Florida in metropolitan statistical areas not connected to septic tanks with demographic and economic factors. Data are from the Florida Department of Health septic tank inspection

1115 permits reported in 2012. Each point represents a county subdivision. Only metropolitan county

1116 subdivisions with at least 5 households and 20 population size were included.

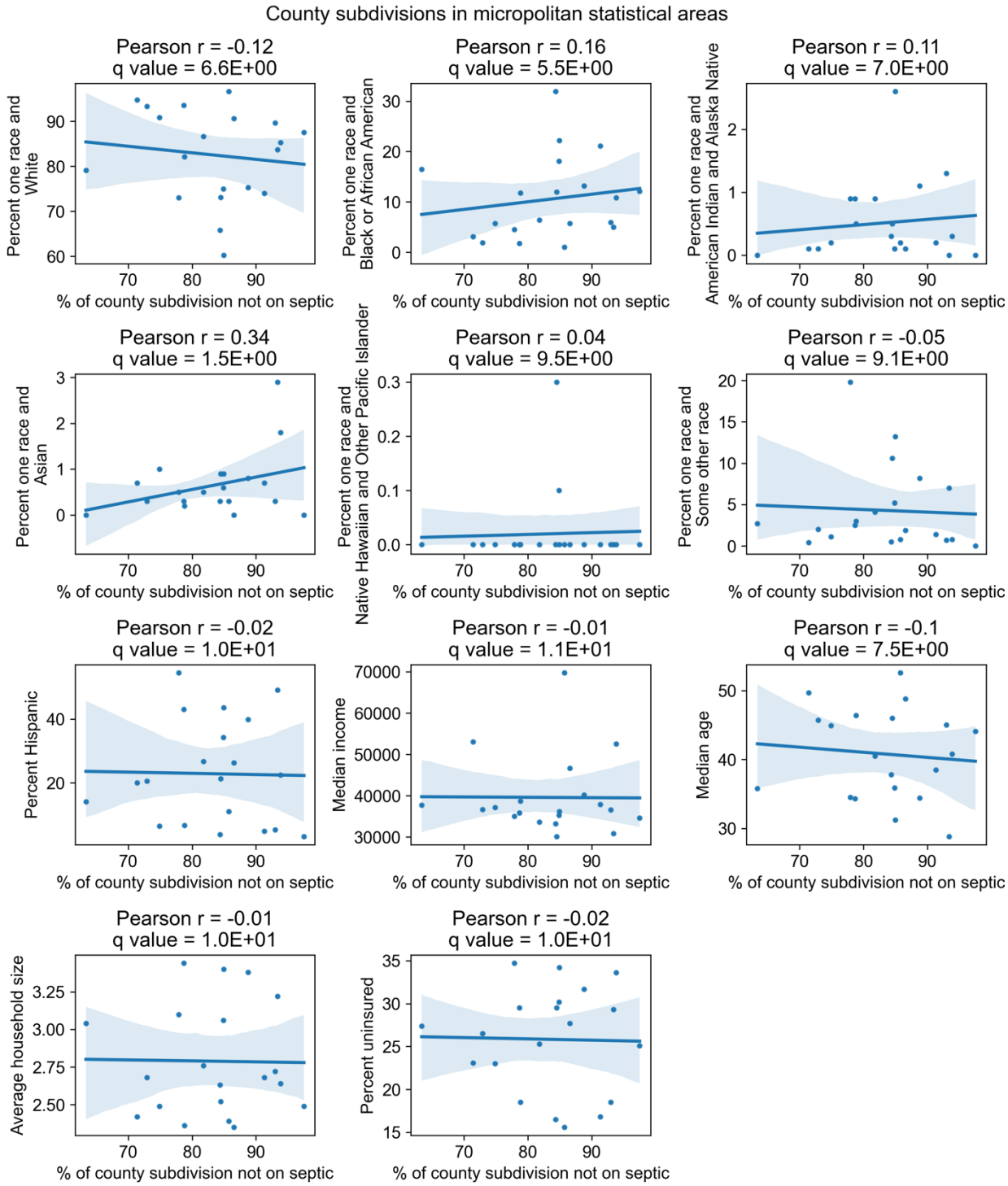


Fig S 7: Correlation of percentage of households in county subdivisions of Florida in micropolitan statistical areas not connected to septic tanks with demographic and economic factors. Each point represents a county subdivision. Data are from the Florida

1121 Department of Health septic tank inspection permits reported in 2012. Only micropolitan county

1122 subdivisions with at least 5 households and 20 population size were included.

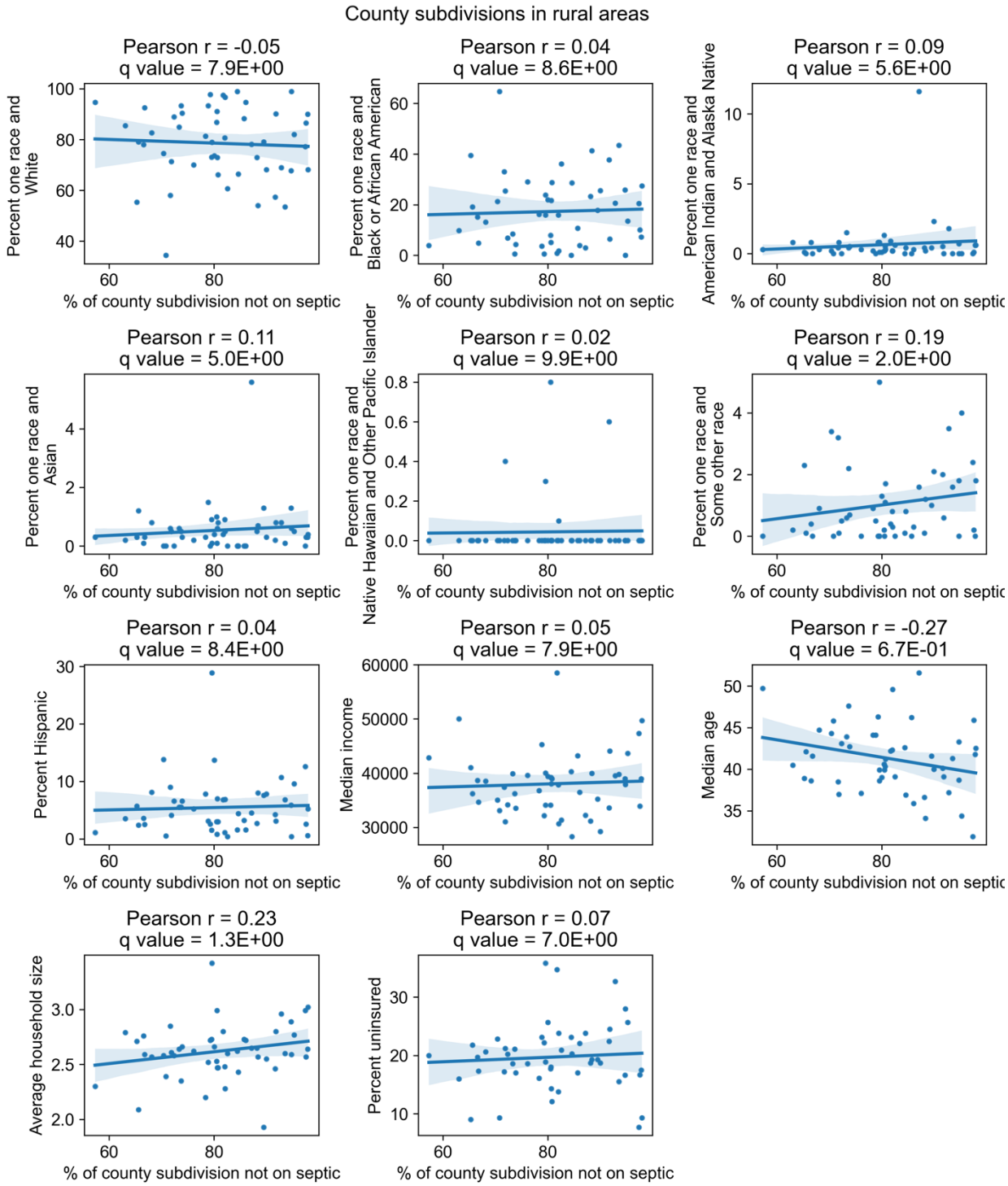


Fig S 8: Correlation of percentage of households in county subdivisions of Florida in rural areas not connected to septic tanks with demographic and economic factors. Data are from the Florida Department of Health septic tank inspection permits reported in 2012. Each

point represents a county subdivision. Only rural county subdivisions with at least 5 households and 20 population size were included.

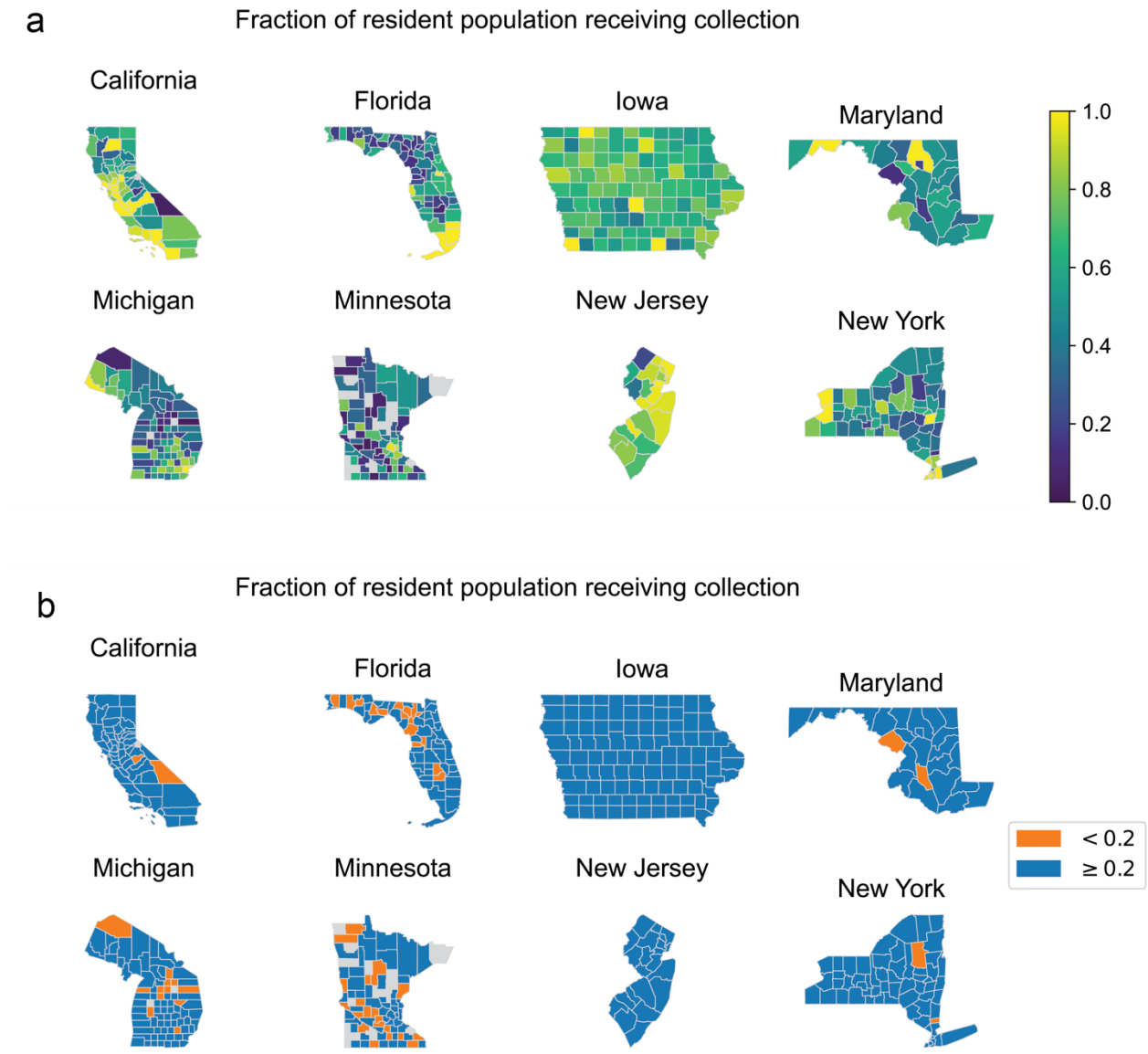


Fig S 9: Fraction of present resident population receiving collection by county in 8 states. Data are from the 2012 EPA Clean Watersheds Needs Survey. (a) Fraction shown on a continuous scale. (b) Fraction above and below 0.2. Only states that had more comprehensive responses in the survey are shown. Counties that did not report any data in the survey are

shown in gray. Counties than report more than 100% of the present resident population receiving collection are shown in (a) as 1 and in (b) as ≥0.2.

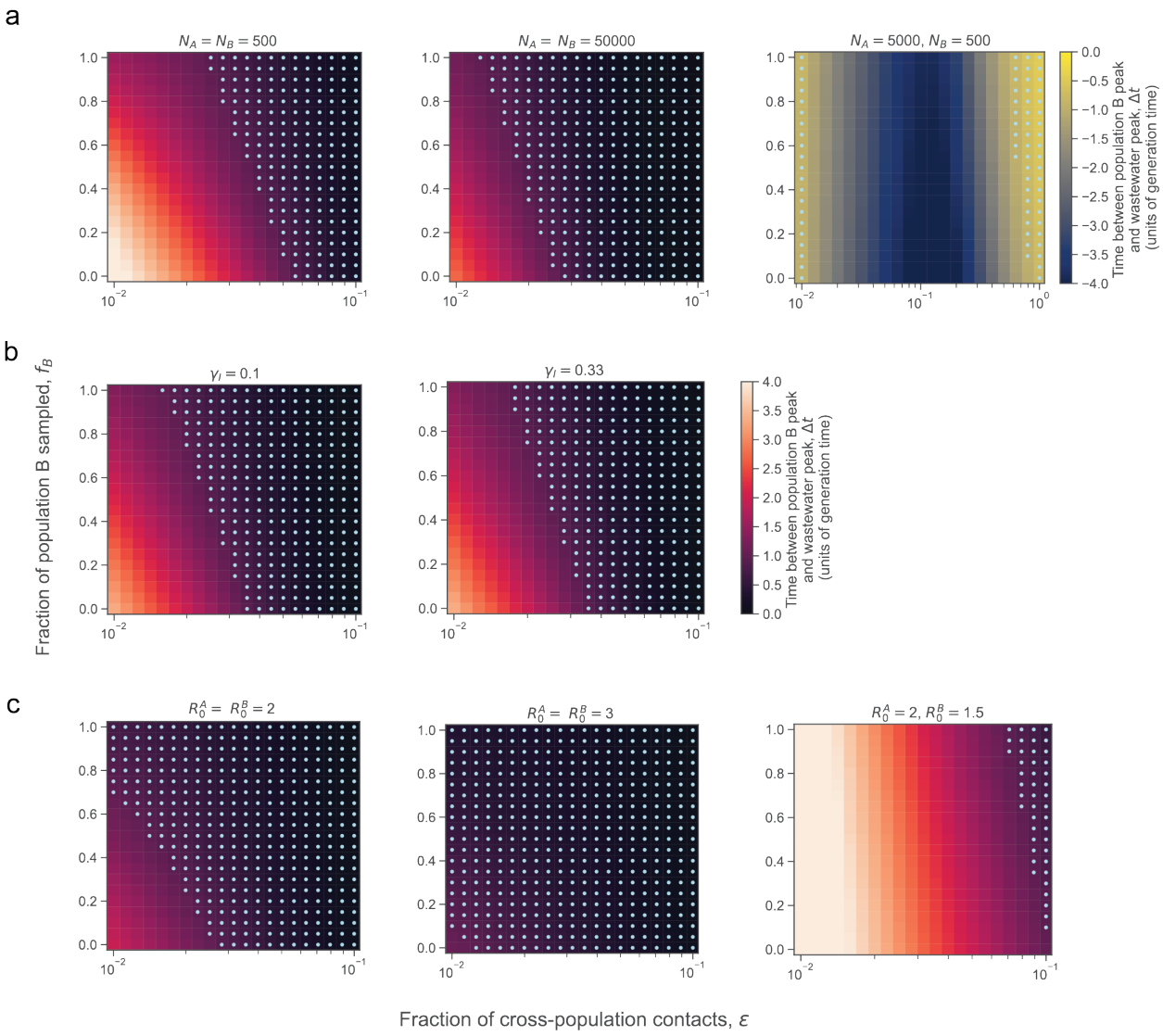


Fig S 10: Sensitivity analysis of Fig 5c from main text. The effect on the time between population B peak and wastewater peak of varying the (a) population size of the two populations (N_A, N_B), (b) recovery rate (γ_I in units of inverse days), and (c) basic reproduction number of the two populations (R_0^A, R_0^B) across a range of values. Unless where indicated, the parameters were N_A=N_B=5000, γ_I=0.18 days^{-1}, and R_0^A, R_0^B=1.5.

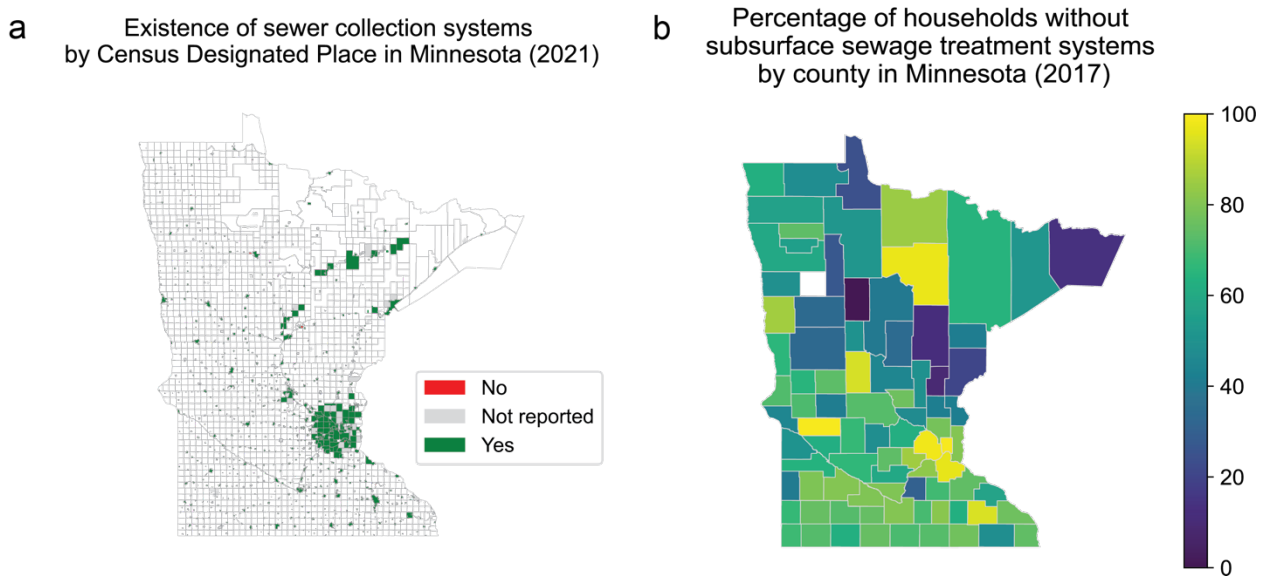


Fig S 11: Sewer connectivity in Minnesota. (a) Map showing whether a community in Minnesota (mapped to Census Designated Place) reported having a collection system in the 2021 Minnesota Wastewater Infrastructure Needs Survey. Almost all communities mapped to Census Designated Places. (b) Percentage of households without subsurface sewage treatment systems (i.e. without septic tanks, suggesting sewered) by county reported in the 2017 Subsurface Sewage Treatment Systems in Minnesota Annual Report [49]. Counties reporting more than 100% of households having subsurface sewage treatment systems are set at 100% (shown as 0% without subsurface sewage treatment systems in the map).

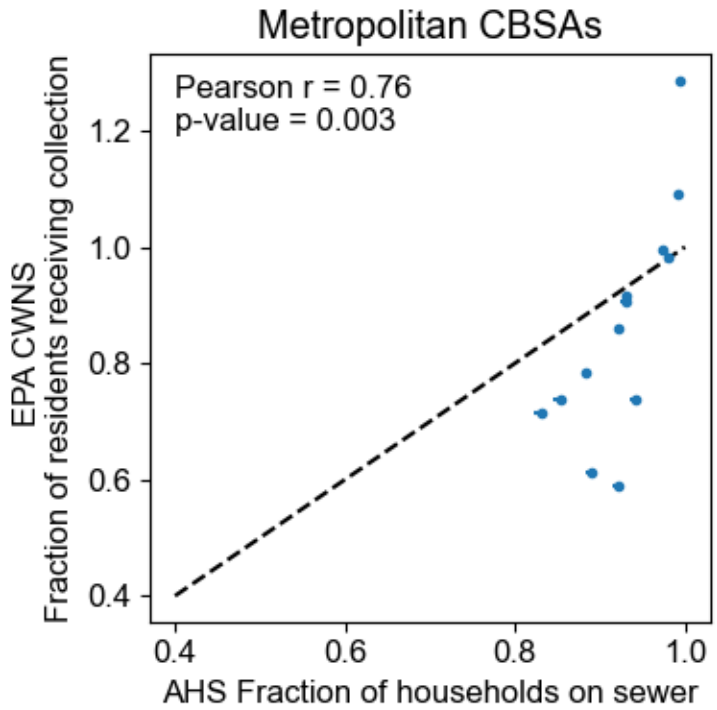


Fig S 12*:* Comparison of the 2012 EPA Clean Watersheds Needs Survey (CWNS) and the 2019 and 2021 U.S. Census American Housing Survey (AHS) datasets in core based statistical areas (CBSA). Data from the EPA CWNS are of the fraction of residents in the CBSA receiving sewage collection. Data from the AHS are of the fraction of households in the CBSA on sewer. Only CBSAs that were oversampled in the AHS data and were in states that had more comprehensive responses in the EPA CWNS were included (see Methods). The dashed black line shows y=x as a reference.

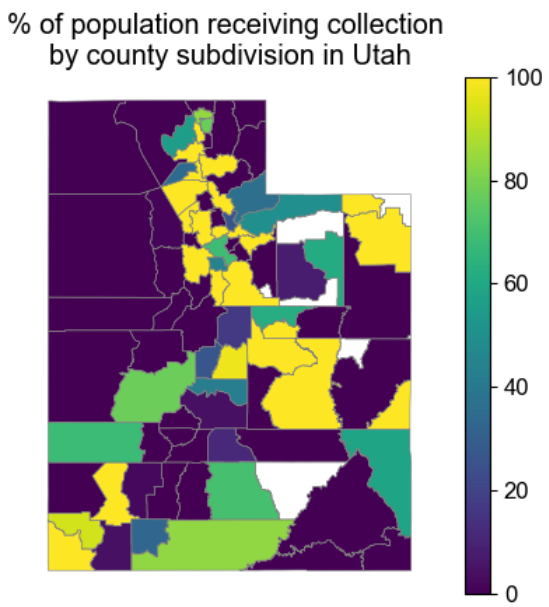


Fig S 13: The percentage of the population that received wastewater collection by a municipal utility in Utah by county subdivision of Utah. Data are from the 2021 Utah Municipal Wastewater Planning Survey. All values less than 0 are displayed as 0. County subdivisions with fewer than 5 households or 20 population size are displayed in white. Note that Indian reservations were not surveyed in this dataset.

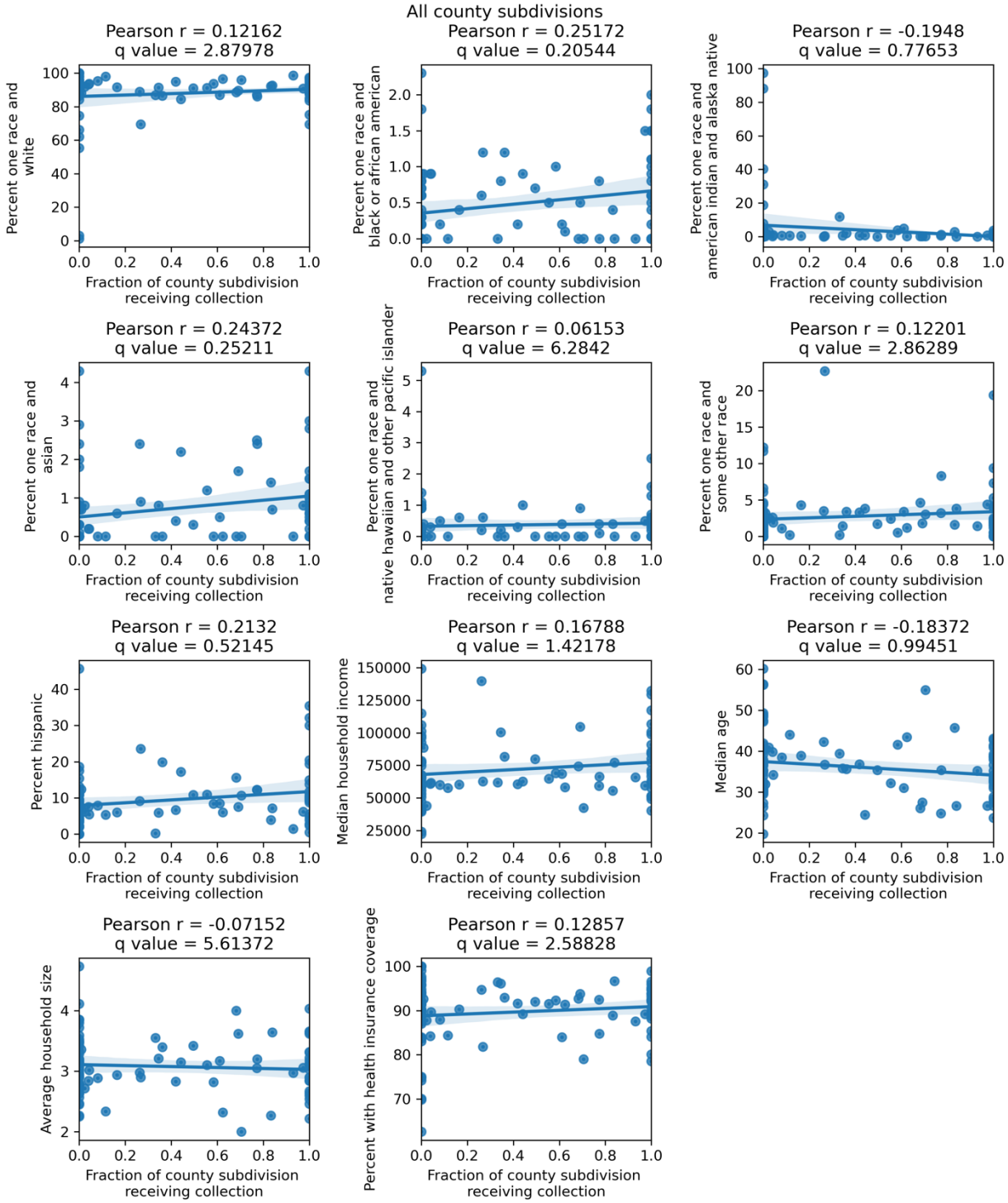


Fig S 14: Correlation of fraction households in county subdivisions of Utah with wastewater collection with demographic and economic factors. Data are from the 2021 Utah Municipal Wastewater Planning Survey. Note that Indian reservations were not surveyed

in this dataset. Values of the fraction of the county subdivision receiving collection greater than 1 were set to 1.

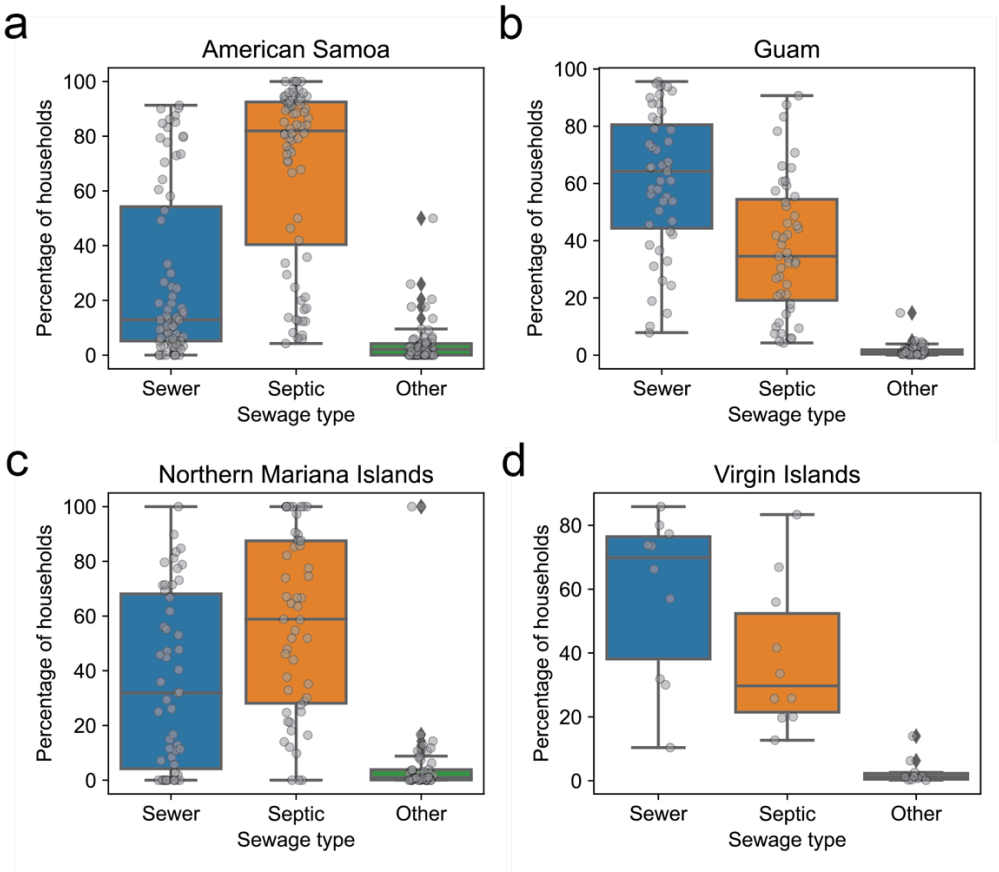


Fig S 15: Fraction of households by Census Designated Place that are connected to sewers, septic tanks, or other forms of sewage disposal in the U.S. Island Areas. (a) American Samoa, (b) Guam, (c) the Northern Mariana Islands, and (d) the Virgin Islands. Data are from the 2020 U.S. Census Island Areas Decennial Survey.

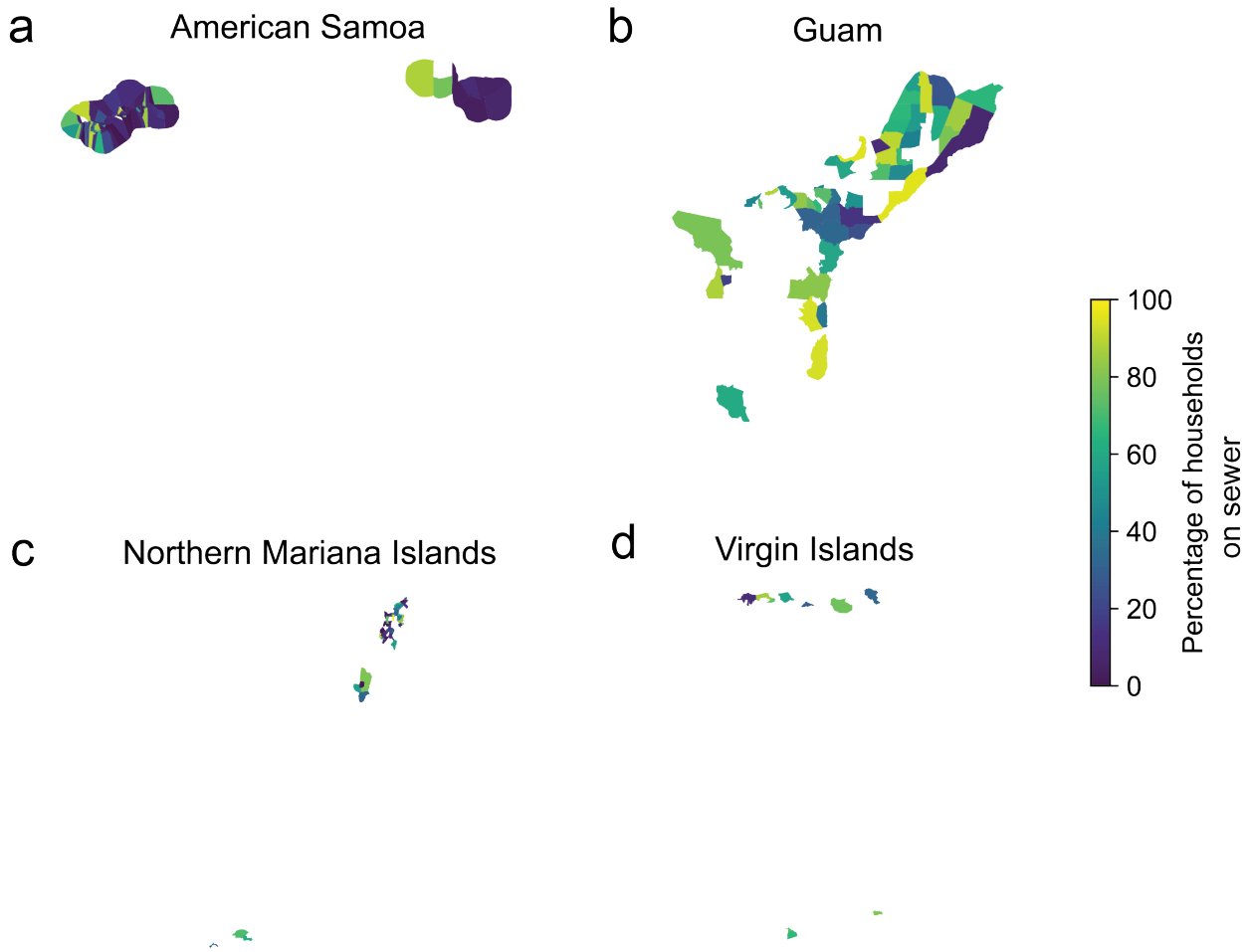


Fig S 16: Map of fraction of households connected to sewers by Census Designated Place. (a) American Samoa, (b) Guam, (c) the Northern Mariana Islands, and (d) the Virgin Islands. Data are from the 2020 U.S. Census Island Areas Decennial Survey.

1187 # Supplementary tables

1188 **Table S 1: Counties in the US with >=5% of occupied housing units lacking complete**

1189 **plumbing facilities.** Data are from the 2021 U.S. Census American Community Survey.

	County, State	Percent of occupied housing units lacking complete plumbing facilities
0	Yukon-Koyukuk Census Area, Alaska	36.1
1	Bethel Census Area, Alaska	31.1
2	Kusilvak Census Area, Alaska	25.3
3	Nome Census Area, Alaska	24.5
4	Northwest Arctic Borough, Alaska	20
5	Denali Borough, Alaska	18.1
6	Lake and Peninsula Borough, Alaska	13.8
7	Apache County, Arizona	12.8
8	Dillingham Census Area, Alaska	12
9	Southeast Fairbanks Census Area, Alaska	11.1
10	North Slope Borough, Alaska	10.8
11	Haines Borough, Alaska	10.5
12	Copper River Census Area, Alaska	10.1
13	San Juan County, Utah	10
14	Adams County, Indiana	9.7
15	McKinley County, New Mexico	9.6

16	Oglala Lakota County, South Dakota	9.2
17	Hoonah-Angoon Census Area, Alaska	7.6
18	Mora County, New Mexico	7.6
19	Navajo County, Arizona	7
20	Prince of Wales-Hyder Census Area, Alaska	6.7
21	Lawrence County, Mississippi	6.2
22	Catahoula Parish, Louisiana	6.1
23	Hudspeth County, Texas	6.1
24	Fairbanks North Star Borough, Alaska	5.6
25	Salinas Municipio, Puerto Rico	5.1
26	Yakutat City and Borough, Alaska	5.1

Table S 2: Correlation of the fraction of households by county connected to public sewers and different demographic or economic factors, stratified by state and Metropolitan, Micropolitan, and Rural Statistical Areas. Only correlations that are significant (q-value<=0.05) are shown.

Demographic or economic variable	State	Metropolitan, Micropolitan, or Rural Statistical Area	Pearson r	q-value
Median income	FL	Micropolitan Statistical Area	0.98	9.8E-03
Percent Hispanic	FL	Metropolitan Statistical Area	0.53	1.8E-02
Percent one race and Asian	MN	Metropolitan Statistical Area	0.72	2.3E-03
	FL	Metropolitan Statistical Area	0.52	3.0E-02
	MN	Metropolitan Statistical Area	0.73	1.1E-03

Percent one race and Black or African American	MI	Metropolitan Statistical Area	0.67	1.4E-02
	MN	Metropolitan Statistical Area	-0.74	7.6E-04
Percent one race and White	MI	Metropolitan Statistical Area	-0.71	4.6E-03
	IA	Metropolitan Statistical Area	-0.69	5.0E-02

Table S 3: Additional datasets discussed in the Supplementary Information but excluded from the main text due to data incompleteness and potential biases.

Dataset	Location	Year(s)	Description	Sampling method	Dataset completeness and potential biases	Ref.
Utah Municipal Wastewater Planning Survey	Utah	2021	This survey reported the population receiving collection by each municipal utility that owns or operates a sanitary sewerage system.	The survey completed was by treatment plant operators. Participation was reported in the survey information as being mandatory.	Of those contacted, 71% responded to the survey [47]. In some county subdivisions, >100% of the population appeared connected to sewers, likely due to errors in estimation, differences in how combined treatment and collection facilities reported population estimates, or collection populations residing in neighboring county subdivisions.	[47]
Minnesota Wastewater Infrastructure Needs Survey (WINS)	Minnesota	2021	Communities (places and townships) responsed to the question "does your community have a collection system?" (yes/no/did not respond).	This was a voluntary survey.	Some communities (number unknown) did not respond to the survey.	[48]
Minnesota Subsurface Sewage Treatment Systems (SSTS)	Minnesota	2017	This dataset reports the number of subsurface sewage treatment systems (SSTS) in various local government units.	Surveys were sent to local government units with a known SSTS program. According to the SSTS Annual Report [49], 'some of the data is "hard" data, such as the reported number and types of permits issued. Other data is considered "soft" data, such as the reported best estimates provided to determine SSTS	Two hundred and eleven out of 218 known SSTS programs submitted an annual report. Not all SSTS programs were identified and contacts were not always provided. Fourteen reported having zero SSTS within their jurisdiction despite permitting SSTS in 2017.	[49]

				compliance rates and, sometimes, the total number of SSTS in each jurisdiction.'		

Table S 4: Manual changes community names reported in the Minnesota Wastewater Infrastructure Needs Survey to be able to match with a U.S. Census Designated Place.

Original name in dataset	**Edited name for merging with Census Designated Place name**	**Note**
St X (i.e. St Paul)	St. X (i.e. St. Paul)	
Plainview - Elgin Sanitary District	[Deleted]	There was another entry for Plainview (same zip code) with the same response for whether there was collection facility
Clear Lake - Clearwater Sewer Authority	Clear Lake	
Tower-Breitung Wastewater Board	Tower	
Birchwood	Birchwood Village	
Coleraine, Bovey, Taconite Joint Wastewater Commission	[Deleted]	Two entries for Coleraine with same zip code and similar population sizes
Glacial Lakes Sanitary Sewer and Water District	[Deleted]	Two entries for Spicer with same zip code and similar population sizes

Central Iron Range Sanitary Sewer District	[Deleted]	Two entries for Chisholm with same zip code and similar population sizes
Shorewood Park Sanitary District	[Deleted]	Two entries for Rush City with same zip code. Deleted the entry with the smaller population size.
Rich Prairie Water & Sewer District (Pierz/Genola)	Pierz	
Pine River Area Sanitary District	Pine River	
Moose Lake Windemere Area Sanitary Sewer District	[Deleted]	Two entries for Moose Lake with same zip code. Deleted the entry with no reported population
Alexandria Lakes Area Sanitary District	Alexandria	
North Long Lake Sewer District	Brainerd	
Serpent Lake Sanitary Sewer District	Crosby	
Blomkest/Svea Sewer Board	Blomkest	
GEM Sanitary District	Greenwald	
Chisago Lakes Joint Sewage Treatment Commission	Center City	

Duluth	[Deleted]	3 entries in Duluth - kept the one with the largest population size (Western Lake Superior Sanitary District (WLSSD))
Duluth – North Shore Sanitary District	[Deleted]	3 entries in Duluth - kept the one with the largest population size (Western Lake Superior Sanitary District (WLSSD))
MCES - Metro Council Env. Services	[Deleted]	Two entries in St. Paul. Deleted the entry with the smaller population size.