

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

3

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

5

6 **Authors:** QinQin Yu<sup>1</sup>, Scott W. Olesen<sup>2</sup>, Claire Duvallet<sup>2</sup>, Yonatan H. Grad<sup>1\*</sup>

7

8 **Affiliations:**

9 1. Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of  
10 Public Health, Boston, Massachusetts, USA

11 2. Biobot Analytics, Inc., Cambridge, Massachusetts, USA

12

13 \*Corresponding author

14 Email: [ygrad@hsph.harvard.edu](mailto:ygrad@hsph.harvard.edu) (YG)

## 15 **Abstract**

16 Wastewater-based epidemiology is a promising public health tool that can yield a more  
17 representative view of the population than case reporting. However, only about 80% of  
18 the U.S. population is connected to public sewers, and the characteristics of populations  
19 missed by wastewater-based epidemiology are unclear. To address this gap, we used  
20 publicly available datasets to assess sewer connectivity in the U.S. by location,  
21 demographic groups, and economic groups. Data from the U.S. Census' American  
22 Housing Survey revealed that sewer connectivity was lower than average when the  
23 head of household was American Indian and Alaskan Native, White, non-Hispanic,  
24 older, and for larger households and those with higher income, but smaller geographic  
25 scales revealed local variations from this national connectivity pattern. For example,  
26 data from the U.S. Environmental Protection Agency showed that sewer connectivity  
27 was positively correlated with income in Minnesota, Florida, and California. Data from  
28 the U.S. Census' American Community Survey and Environmental Protection Agency  
29 also revealed geographic areas with low sewer connectivity, such as Alaska, the Navajo  
30 Nation, Minnesota, Michigan, and Florida. However, with the exception of the U.S.  
31 Census data, there were inconsistencies across datasets. Using mathematical modeling  
32 to assess the impact of wastewater sampling inequities on inferences about epidemic  
33 trajectory at a local scale, we found that in some situations, even weak connections  
34 between communities may allow wastewater monitoring in one community to serve as a  
35 reliable proxy for an interacting community with no wastewater monitoring, when cases  
36 are widespread. A systematic, rigorous assessment of sewer connectivity will be

- 37 important for ensuring an equitable and informed implementation of wastewater-based
- 38 epidemiology as a public health monitoring system.

## 39 Introduction

40 Wastewater-based epidemiology (WBE) plays an important role in surveillance of  
41 SARS-CoV-2 [1,2], polio virus [3,4] and other pathogens [5,6] and has applications to  
42 monitoring a variety of other public health concerns [7], including opioid usage [8]. One  
43 proposed benefit of wastewater-based epidemiology is that wastewater data is more  
44 representative of the population than case reporting, which can be biased towards those  
45 with health-seeking behavior or access to healthcare [9]. For example, the populations  
46 served by the Health and Human Services SARS-CoV-2 National Wastewater  
47 Surveillance System were more representative of the entire US's age distribution and  
48 Black and Hispanic populations than the vaccinated population [10].

49  
50 While WBE offers convenient sampling of populations served by public sewers, about  
51 20% of individuals in the US live in homes not connected to public sewers [11,12]. This  
52 includes those on decentralized wastewater systems and those with no wastewater  
53 treatment systems. The most common decentralized wastewater system is septic tanks  
54 [13], which collect and treat wastewater onsite, typically the yard of the home.  
55 Households without wastewater treatment systems may have outhouses or privies,  
56 chemical toilets, or no plumbing. Variability in sewer connectivity exists across the U.S.  
57 [14,15], with more connectivity in urban areas than rural areas [11,16–18] and  
58 disparities driven by structural inequalities [19].

59  
60 As with any emerging public health tools, it is important to ask to what extent the tool  
61 exacerbates or alleviates inequities. A few studies have evaluated equity of sewer

62 connectivity on broad geographic scales or at the sewershed level. A 2017 study from  
63 Environmental Protection Agency Office of Water showed that households in the U.S.  
64 that earned less than the national median household income (MHI, \$61,000) were  
65 almost 10% more likely to have a decentralized wastewater system or no wastewater  
66 treatment system compared to households that earned more than the MHI [20].  
67 Additionally, this study found that as household income decreased, decentralized  
68 wastewater system usage increased in Florida, Hawaii, and Delaware, but not in Rhode  
69 Island. Based on the 2019 U.S. Census Bureau's American Housing Survey, rural areas  
70 are less connected to sewers, and while the income in rural areas was lower than that in  
71 metropolitan areas, households connected to septic tanks were wealthier than those  
72 connected to sewers within both rural and metropolitan areas [11]. In sewersheds in  
73 North Carolina with comprehensive sewershed maps, the sewered population had  
74 higher social vulnerability, more minorities, lower income, and lower educational  
75 attainment than the unsewered population [21].

76  
77 However, we have an incomplete understanding of the factors associated with sewer  
78 connectivity across the US and the implications for the interpretation of wastewater  
79 data. In this study, we sought to address the following questions, focusing our analyses  
80 on the U.S.: (1) To what extent is there demographic and economic inequity in sewer  
81 connectivity? (2) Which geographic areas have low sewer connectivity? (3) What is the  
82 applicability of WBE data to neighboring unsampled communities? To address the first  
83 question, we analyzed household-level data of sewer connectivity stratified by  
84 geographic, demographic, and economic variables. For the second question, we

85 evaluated datasets aggregated at the county or county subdivision levels to qualitatively  
86 identify geographic areas that have low sewer connectivity. To address the third  
87 question, we used a mathematical model to simulate WBE in two interacting  
88 populations.

89

## 90 **Methods**

### 91 **Dataset compilation**

92 We assembled publicly available datasets on sewer connectivity, septic connectivity, or  
93 no plumbing at the household level across the U.S. from federal and state agencies  
94 informed by (1) discussions with experts from federal and state agencies and (2) web  
95 searches using terms including “Number of people connected with treatment plants in  
96 United States”, “Number of households on septic”, and “U.S. census sewer”. The types  
97 of data included household-level data, locations of sewer systems and population  
98 served, and population served by or lacking wastewater systems aggregated by county  
99 or county subdivision. All datasets are summarized in **Table 1** and described in more  
100 detail in the Supplementary Information.

101

102 **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	The sampling method was a representative sample of all U.S. households at the time of sample selection (most	The survey excludes group quarters, businesses, hotels, and motels. Additionally, the survey was more likely to classify rural seasonal homes as vacant units, counted	[12,22–26]

			characteristics of approximately 60,000-80,000 representative households.	recently in 2015 and before that, in 1985). New housing units were added each survey cycle.	large cluster septic systems as public sewer, and was thought to lose new households built in rural areas due to its longitudinal design.	
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.	[27]
Florida Department of Health septic tank inspections	Florida	Up to 2012	This dataset includes the onsite sewage treatment and disposal systems locations inspected by the Florida Department of Health reported in June 2012.	The sample included onsite sewage treatment and disposal system permits recorded in the Florida Department of Health system.	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 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 [28].	[29]
EPA Clean Watersheds Needs Survey	US	2012	This survey included voluntary submissions of the locations of publicly owned wastewater collection and treatment facilities	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	[30,31]

			and their estimated population served for the purposes to assess funding needs for treatment works projects.		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.	
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 in person and via phone interviews.	The COVID-19 pandemic impacted collection of data on group quarters [32] and the survey had higher nonresponse rates than previous years' surveys, particularly for Guam. Quality control checks were performed on responses.	[33]

103 The datasets are further described in the S1 Appendix and the dataset completeness  
 104 and potential biases are further assessed in the S2 Appendix. Additional datasets that  
 105 are briefly described in the main text but primarily analyzed in the Supplementary  
 106 information are summarized in **Table S 4**.

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

109 All summary statistics were calculated by applying the weights reported in the AHS Public Use  
 110 File [26]. Because not every household was sampled and there was uneven sampling of  
 111 households across the U.S., the weight estimates the number of similar households that each  
 112 surveyed household represented. Error bars on summary statistics represent the middle 95% of  
 113 values calculated using the reported replicate weights (160 replicate weights per household).



114 Each summary statistic was calculated for each replicate separately before taking the middle  
115 95% of values. Only categories with at least 5 households were used.

116

## 117 **Analysis of sewer connectivity by urban or rural areas in**

### 118 **AHS data**

119 We used the most recent AHS National Sample with publicly available data on the  
120 urban or rural location of the household (2013) which categorized households into being  
121 in a central city of a Metropolitan Statistical Area (MSA), in an urban area of an MSA but  
122 not in a central city, in a rural area of an MSA but not in a central city, in an urban area  
123 outside of an MSA, or in a rural area outside of an MSA [34]. Central cities are defined  
124 as having either (1)  $\geq 250,000$  population or at least 100,000 people working within  
125 corporate limits, (2)  $\geq 25,000$  population, at least 75 jobs for each 100 residents who  
126 were employed, and 60% or fewer of the city's resident workers commuted to jobs  
127 outside, or (3) 15,000-25,000 population, at least one third the size of the metropolitan  
128 statistical area's largest city, and met the two commuting requirements in (2).

129 Metropolitan Statistical Areas were defined as whole counties that have significant  
130 levels of community and contiguous urban areas in common and may include rural  
131 areas. Urban areas were defined as having  $\geq 2500$  people, with at least 1500 residing  
132 outside of institutional group quarters. Rural areas were defined as those not in urban  
133 areas.

134

## 135 **Analyses of geographic areas with low connectivity to sewer**

136 Data of individual locations of septic tanks or sewer collection systems, when they  
137 existed, were aggregated based on latitude and longitude into county or county  
138 subdivision boundaries as defined by the U.S. Census' TIGER shapefiles of the  
139 corresponding year [35]. To calculate the fraction of a geographic area connected to  
140 septic or sewer, the total population size or total number of households in that  
141 geographic area was taken from the corresponding year in the American Community  
142 Survey (ACS) 5-year estimate [27]. For county subdivision and census designated place  
143 analyses, only areas with at least 5 households and 20 population size were used for  
144 the analyses. This thresholding removed at most 1% of the total population and at most  
145 0.03% of the total number of households, making it unlikely to bias our results. Detailed  
146 data analyses of the state and island area datasets are described in the Supplementary  
147 information.

148  
149 Counties and county subdivisions were classified into Metropolitan statistical areas (at  
150 least one urbanized area of 50,000 or more population), Micropolitan statistical areas (at  
151 least one urban cluster of 10,000-50,000 population), or Rural areas (all other areas)  
152 using the U.S. Census classification of the encompassing Core Based Statistical Area  
153 (core area containing a substantial population nucleus together with adjacent territory  
154 that has a high degree of social and economic integration with the core as measured by  
155 commuting ties) [36].

156  
157 **Demographic and economic variable correlations with sewer**  
158 **or septic connectivity**

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

168

## 169 **Simulations of interacting populations and wastewater**

### 170 **sampling**

171 Two interacting populations (A and B) were modeled in a deterministic compartmental model  
172 with susceptible, infected, and recovered (SIR) compartments for each population and a  
173 fraction,  $\varepsilon$ , of cross-population contacts. The base parameters were set at population size  
174  $N_A=N_B=5000$ , recovery rate  $\gamma=0.18$  inverse days (corresponding to an infectious period of 5.6  
175 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  
176 analyses. The differential equations were solved using the integrate.odeint function in the scipy  
177 package (version 1.6.2) in Python (version 3.11.0) using a timestep of 0.1 days.

178

179 Wastewater sampling was modeled by the number of copies of pathogen genetic material shed  
180 per day by each population that was sampled by wastewater divided by the volume of sampled  
181 wastewater produced per day by each population. Both numerator and denominator depended  
182 on the fraction of waste produced (including shed pathogen genetic material) by each

183 population that is sampled by wastewater,  $f_A$  and  $f_B$ . A detailed description of the simulations is  
184 provided in S1 Appendix.

185

## 186 **Code availability**

187 All code and data to reproduce the analyses and simulations are available at

188 [https://github.com/gradlab/wastewater\\_equity](https://github.com/gradlab/wastewater_equity).

189

## 190 **Results**

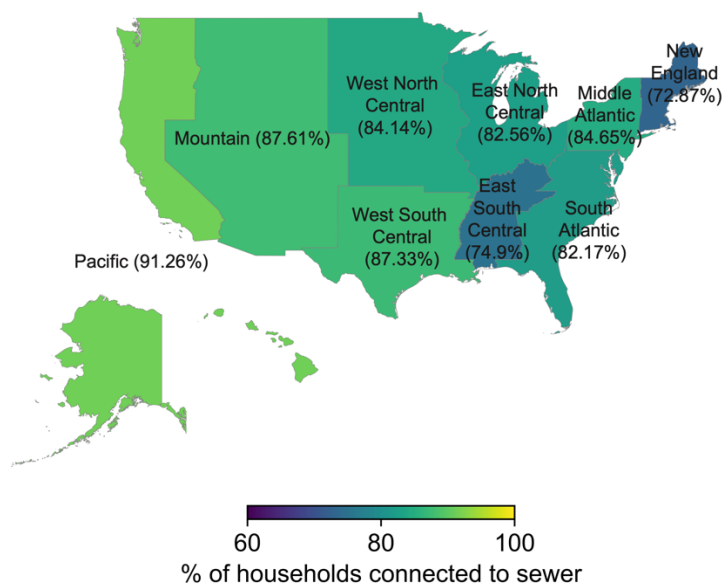
### 191 **Is there inequity in who could be sampled by wastewater-** 192 **based epidemiology?**

193 To address this question, we used the American Housing Survey (AHS), a biannual  
194 longitudinal survey of representative housing units within the U.S that reports  
195 household-level data of sewer connectivity, demographic factors, and economic factors  
196 (**Table 1**). We focused our analyses on geographic scales with a reliable sample size for  
197 statistical analyses (census division, urban/rural at the national level, metropolitan  
198 statistical areas) [37].

199

200 According to the 2021 AHS, 83% of US households were connected to public sewers  
201 and 16% to standard septic tanks (**Fig S 1**). Each of the other forms of sewage disposal  
202 (non-standard septic tanks, chemical toilets, outhouses or privies, other, none, or not  
203 reported) was more than an order of magnitude less common (**Fig S 1**). Across census  
204 divisions in 2021, the western US was overall better connected to sewers than the

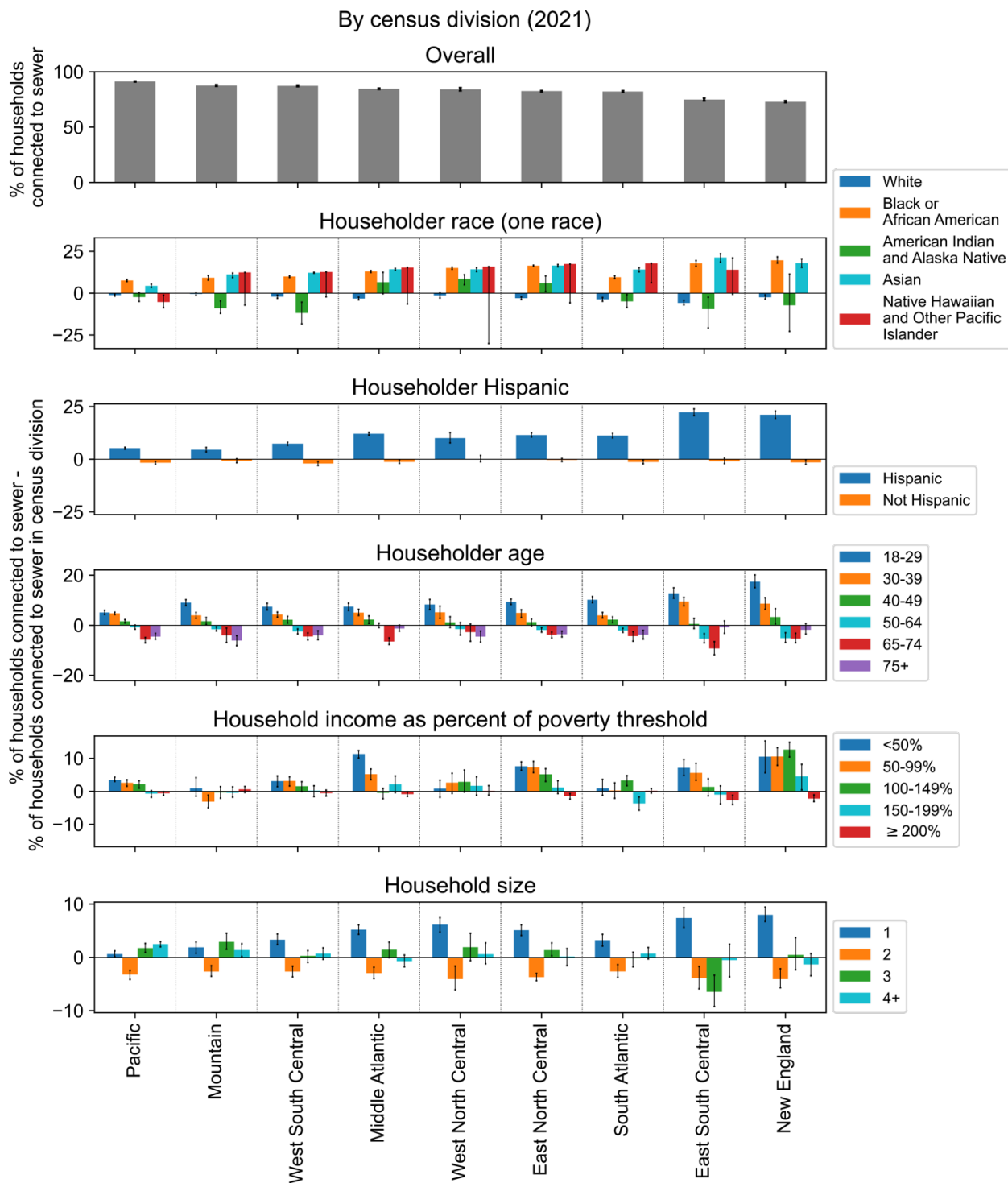
205 eastern US, with the highest levels of sewer connection in the Pacific (91.3% of  
206 households, CI: 90.8%-91.7%) and Mountain (87.6% of households, CI: 86.8%-88.4%)  
207 divisions and the lowest levels of sewer connection in the New England (72.9% of  
208 households, CI:72.1%-74.0%) and East South Central (74.9% of households, CI:  
209 73.7%-76.3%) divisions (**Fig 1**).



210  
211 **Fig 1: The percentage of households connected to sewers in each census division.** Data  
212 are from the 2021 American Housing Survey National Survey [12]. The number of sampled  
213 households for each census division are 3890 (New England), 7018 (Middle Atlantic), 8703  
214 (East North Central), 2862 (West North Central), 13643 (South Atlantic), 2903 (East South  
215 Central), 8051 (West South Central), 4570 (Mountain), and 12501 (Pacific).  
216  
217 Across census divisions, households with an Asian, Black or African American, or Native  
218 Hawaiian and Other Pacific Islander householder (the owner or renter of the unit) were on  
219 average 14.0%, 13.1%, and 12.5% more connected to sewers, respectively, than the overall

220 census division; however, specifically in the Pacific census division, households with a Native  
221 Hawaiian and Other Pacific Islander householder were 5.3% less connected to sewers than the  
222 overall census division (**Fig 2**). Households with an American Indian and Alaska Native or White  
223 householder were both on average 2.6% less connected to sewers than the overall census  
224 division. Across census divisions, households with a Hispanic householder were on average  
225 11.7% more connected to sewers whereas households with a non-Hispanic householder were  
226 1.1% less connected to sewers than the overall census division.

227  
228 Connection to sewers decreased with the age of the householder (**Fig 2**), with the exception  
229 that households with a householder aged 75+ were better connected in most census divisions  
230 than those with a householder aged 65-74 (79.7% vs 77.9% of households). Connection to  
231 sewers decreased with household income, as measured as a percentage of the federal poverty  
232 threshold determined by the Census [38], with the exception of the Mountain region where  
233 households with an income at 50-99% of the poverty level were less connected than those with  
234 income at  $\geq 200\%$  of the poverty level (84.5% vs 88.2% of households). Additionally, households  
235 with 1 person were on average 4.2% more connected whereas households with 2 people were  
236 on average 3.6% less connected to sewers than the overall census division. There were no  
237 consistent trends across census divisions for households with more than 2 people.



238

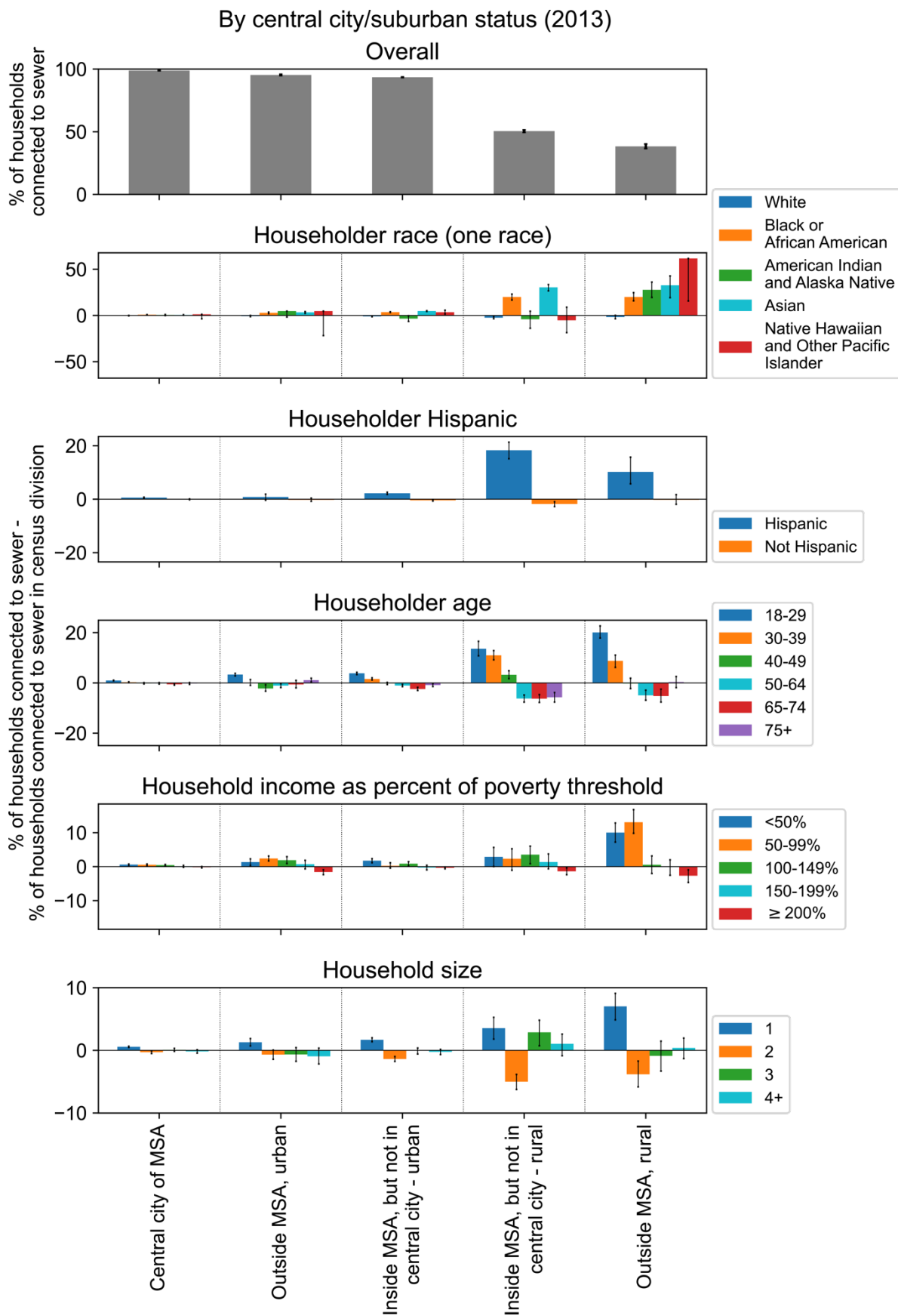
239 **Fig 2: The percentage of households connected to sewers in each census division by**

240 **demographic and economic characteristics. Data are from the 2021 American Housing**

241 Survey National Survey [12]. The number of sampled households for each category is shown in  
242 **Fig S 2**.

243  
244 We sought to describe how the differences in sewer connection between demographic and  
245 economic groups observed at the census division level was affected by household location in  
246 urban or rural areas (see Methods). The qualitative trends observed at the census division level  
247 for race, ethnicity, age, household size, and income were largely preserved across national  
248 urban and rural areas (**Fig 3**). One notable exception was that lower sewer connectivity for  
249 households with an American Indian and Alaska Native or Hawaiian and Other Pacific Islander  
250 householder compared to those with a Black or Asian householder was only observed in  
251 locations inside of an Metropolitan Statistical Area (MSA) but outside of a central city (both  
252 urban and rural). In central cities and locations outside of an MSA (both urban and rural),  
253 households with an American Indian and Alaska Native or Hawaiian and Other Pacific Islander  
254 householder had similar levels of sewer connectivity as households with a Black or Asian  
255 householder.





257 **Fig 3: The percentage of households connected to sewers by urban and rural**  
258 **categorizations and by demographic and economic characteristics.** See Methods for  
259 definitions of Metropolitan Statistical Area (MSA), central city, urban, and rural. Data from the  
260 2013 American Housing Survey National Survey [39]. The number of sampled households for  
261 each category is shown in **Fig S 3**. MSA: metropolitan statistical area.

262

263 Next, we assessed sewer connectivity in large metropolitan areas, which was the  
264 smallest geographic scale resolved by the AHS due to sample size. In 35 of the largest  
265 metropolitan areas in 2019 and 2021, household connectivity to sewers ranges from  
266 69.2% (Birmingham-Hoover, AL) to 99.4% (LA-Long Beach-Anaheim, CA) (**Fig S 4**). In  
267 each metropolitan area, households with a White or American Indian and Alaska Native  
268 householder, with an older householder, with higher income, with a larger household  
269 size, and with a non-Hispanic householder were less connected to sewers than the  
270 average household, which was similar to what was seen when stratifying by census  
271 division. Exceptions were lower connectivity for households with a Hispanic  
272 householder in San Jose, CA (Hispanic householder: 96.1% of households, CI: 95.1%-  
273 97.2%; non-Hispanic householder: 97.8% of households, CI: 97.5%-98.2%) and lower  
274 connectivity for households with lower income in San Francisco, CA; Dallas, TX;  
275 Chicago, IL; and Tampa, FL (**Fig S 4**).

276

277 In summary, at the census division level, broad trends of sewer connectivity by racial  
278 group revealed higher connectivity for households with an Asian, Black or African  
279 American, Native Hawaiian and Other Pacific Islander, or Hispanic householder and  
280 lower connectivity for households with an American Indian and Alaska Native or White

281 householder. Sewer connectivity decreased with age, decreased with household  
282 income, and was highest in households with one person. These trends were similar  
283 when stratifying by urban, rural, central city, and MSA status, except that the broad  
284 racial disparities described above were only observed in urban and rural areas that are  
285 inside of an MSA but outside of a central city. Broad trends were also similar for  
286 individual large metropolitan areas, with some exceptions for households with a  
287 Hispanic householder and trends by household income in particular MSAs.

288

## 289 **Which geographic areas have low connectivity to sewers and** 290 **what are their demographics?**

291 Next, we sought to assess which geographic areas in the U.S. have low connectivity to  
292 sewer systems. These are areas where wastewater-based epidemiology may provide  
293 little direct information on disease dynamics. Additionally, we assessed correlations of  
294 sewer connectivity at the county or county subdivision level with demographic and  
295 economic factors. In the absence of comprehensive national data on sewer connectivity,  
296 we assembled publicly available state-based datasets, which included California,  
297 Florida, Iowa, Maryland, Michigan, Minnesota, New Jersey, and New York, as well as  
298 the U.S. Island Areas. The compilation of the datasets and an assessment of data  
299 completeness are described in the S2 Appendix, **Table 1**, and **Table S 4**.

300

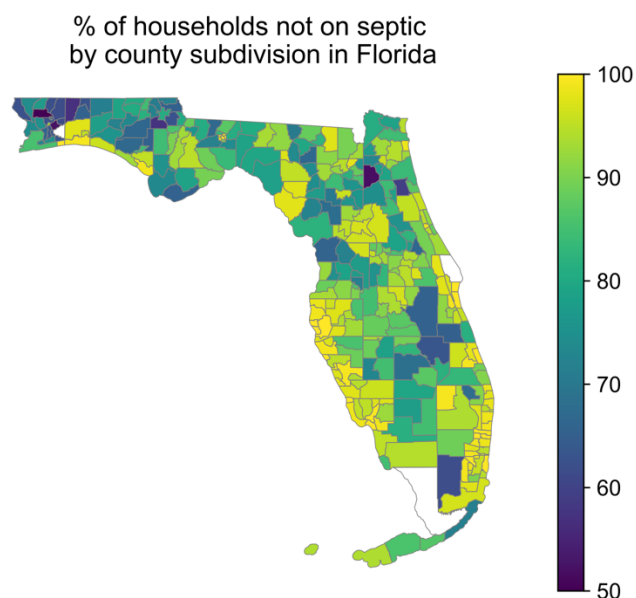
301 Using data from the 2021 American Community Survey on the percentage of occupied  
302 housing units lacking complete plumbing facilities by county, we defined the lower  
303 bound on the percentage of the population not connected to sewers. In 14 counties,

304  $\geq 10\%$  of occupied housing units lacked complete plumbing facilities; of these counties,  
305 12 were in Alaska and 2 were in the Navajo Nation. In 5 counties,  $\geq 20\%$  of occupied  
306 housing units lacked complete plumbing facilities, all in Alaska. The Yukon-Koyukuk  
307 Census Area in Alaska, had the highest percentage of occupied housing units lacking  
308 complete plumbing facilities at 36% (**Fig S 5, Table S 1**).

309  
310 The Florida Department of Health onsite sewage treatment and disposal systems  
311 inspection data (**Fig 4**) revealed that the percentage of households not on septic tanks  
312 (expected to correlate with the percentage of households connected to sewer) was  
313 higher in metropolitan compared to micropolitan county subdivisions and higher in  
314 micropolitan compared to rural county subdivisions, but the effect was not significant ( $p$   
315 = 0.052 and  $p = 0.37$ , respectively) (**Fig S 6a-b**). The percentage of households not on  
316 septic tanks was significantly lower in county subdivisions in the panhandle (74.0% vs  
317 92.0%,  $p < 10^{-9}$ ) and in county subdivisions that did not border the coast (86.6% vs  
318 95.7%,  $p < 10^{-9}$ ) (**Fig S 6**). By combining this dataset with the 2012 American Community  
319 Survey (ACS) describing demographic and economic characteristics of county  
320 subdivisions, we found that county subdivisions with less septic usage (suggesting more  
321 sewer connectivity) were significantly more Hispanic, more Asian, and had a smaller  
322 household size (**Table 2 and Fig S 7**), consistent with the nation-wide results from the  
323 AHS. These trends were driven by metropolitan county subdivisions (**Fig S 8**), and an  
324 additional trend in metropolitan county subdivisions not observed for all county  
325 subdivisions was that county subdivisions with less septic usage had a significantly  
326 lower percentage of American Indian and Alaska Natives. There were too few

327 micropolitan and rural county subdivisions to observe significant trends (**Fig S 9 and Fig**  
 328 **S 10**).

329



330

331 **Fig 4: Map of the percentage of households in each county subdivision of Florida *not***  
 332 **connected to septic tanks.** Data are from the Florida Department of Health septic tank  
 333 inspection data from June 2012. County subdivisions with fewer than 5 households or 20  
 334 population size (<0.01% of total households and <0.01% of total population size in the state) are  
 335 displayed in white.

336

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

Demographic or economic variable	Correlation of variable with % of county subdivision <i>not</i> on septic (Pearson correlation coefficient)	Significance (q-value)
----------------------------------	---	------------------------

Percent one race and Asian	0.31	<b>&lt;0.001</b>
Percent Hispanic	0.21	<b>&lt;0.01</b>
Average household size	-0.19	<b>&lt;0.01</b>
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

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

341  
 342 In the 8 states with robust data in the EPA Clean Watersheds Needs Survey, multiple  
 343 neighboring counties in Florida, Michigan, and Minnesota had less than 20% of  
 344 residents receiving sewage collection, whereas this was not the case in California,  
 345 Florida, Iowa, Maryland, New Jersey, and New York (**Fig S 11**). Across multiple states,  
 346 sewer connectivity by county was positively correlated with percent Asian and percent

347 Black or African American and negatively correlated with percent White (**Table 3**, full  
 348 table shown in **Table S 2**), consistent with the nationwide trends from the AHS data.  
 349 While counties more connected to sewers appeared to have a significantly lower  
 350 percent of American Indian and Alaskan Natives in California, this is affected by missing  
 351 data from Indian reservations. Age and percent without health insurance were  
 352 negatively correlated with sewer connectivity. Counties more connected to sewers had  
 353 significantly higher income in Minnesota, Florida, and California, in contrast to the  
 354 national results from the AHS data. Significant trends of sewer connectivity with  
 355 demographic and economic factors were predominantly driven by metropolitan counties  
 356 (**Table S 3**).

357

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

Correlation of variable with % of households by county connected to public sewers (Pearson correlation coefficient)      Significance (q-value)			
Demographic or economic variable	State	coefficient)	e (q-value)
Median age	MN	-0.43	<0.01
	MI	-0.41	0.01
Median income	MN	0.44	<0.01
	FL	0.48	<0.01
	CA	0.49	<0.01

Percent one race and			
American Indian and			
Alaska Native	CA	-0.50	<0.01
	MN	0.58	<0.001
	MI	0.51	<0.001
	NY	0.55	<0.001
Percent one race and	CA	0.56	<0.001
Asian	FL	0.53	<0.001
	MN	0.62	<0.001
	MI	0.59	<0.001
Percent one race and	NY	0.55	<0.001
Black or African	IA	0.38	<0.01
American	CA	0.46	0.02
	MI	-0.57	<0.001
	CA	-0.54	<0.001
Percent one race and	NY	-0.51	<0.01
White	IA	-0.35	0.03
Percent uninsured			
(health insurance)	MI	-0.38	0.03

360 Only correlations that are significant ( $q\text{-value} \leq 0.05$ ) are shown. Data are from the 2012

361 EPA Clean Watersheds Needs Survey.

362



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

366

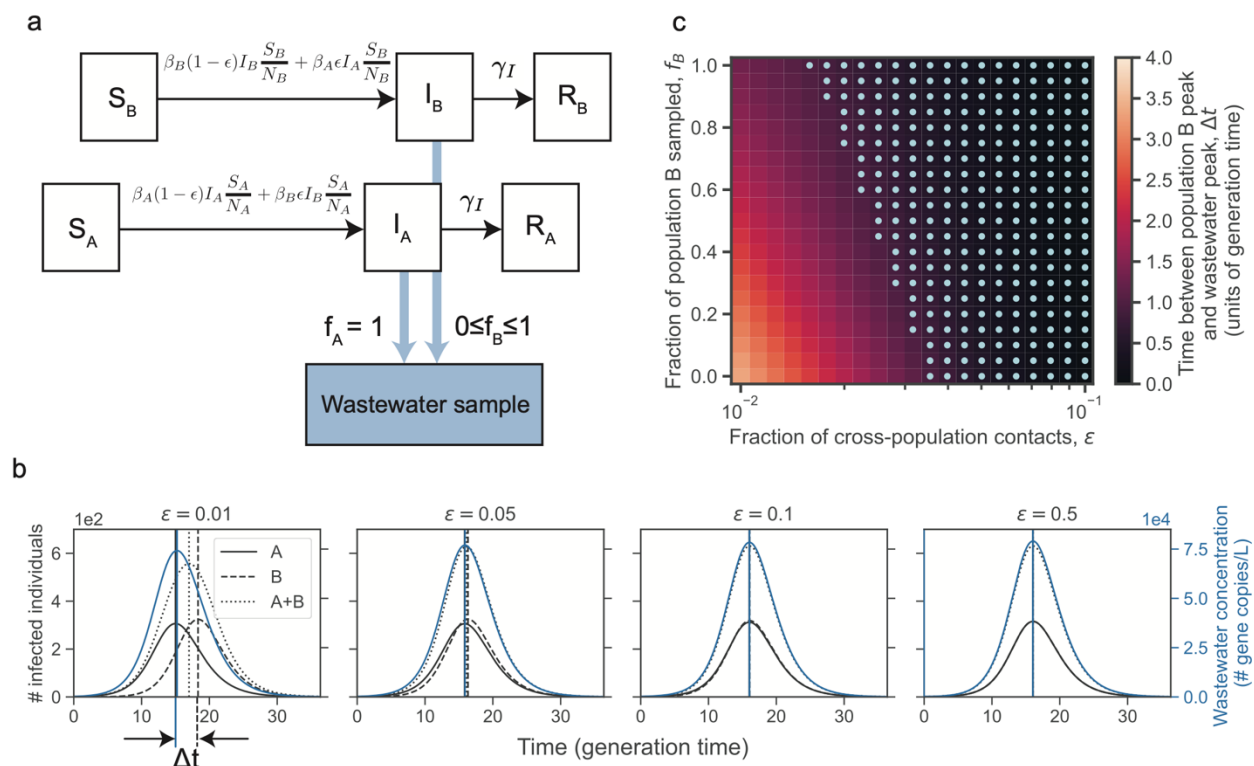
## 367 **Factors influencing applicability of wastewater-based** 368 **epidemiology to communities lacking sewer connection**

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

377

378 A common application of wastewater data is to aid in determining when an outbreak has  
379 peaked, which can inform policy decisions on when to ease restrictions. We first asked  
380 how well the wastewater data could predict the time of peak infections in population B  
381 when it is completely unconnected to sewer systems and thus the only sampling is from  
382 population A. We found that when the two populations have similar sizes, the  
383 wastewater concentration peak and the infection peak in population B was within one  
384 generation time except when the interactions between the two population is weak (**Fig**  
385 **5b**). The generation time is the average time between an individual's infection and

386 transmission, which in an SIR model is the same as the infectious period. For example,  
 387 in two populations with 5000 individuals each, a disease basic reproduction number of  
 388 1.5, and infectious period of 5.6 days, the percentage of contacts that occur across  
 389 populations must drop below 4% before the peak in wastewater concentration occurred  
 390 more than one generation time apart from the peak in infections in the unconnected  
 391 population (Fig 5c).



392  
 393 **Fig 5: Deterministic compartmental model of two interacting populations with**  
 394 **susceptible, infected, and recovered compartments and sampling by wastewater. (a)**  
 395 Schematic of model. Population A is entirely sampled by wastewater and a fraction of  
 396 population B,  $f_B$ , is sampled by wastewater. (b) Number of infected individuals (black curves) in  
 397 population A, population B, and population A and B combined and wastewater concentration  
 398 (blue curves) over time for different interaction strengths or fractions of cross-population  
 399 contacts ( $\epsilon$ ). The vertical lines indicate the peak time for the curves with matching line style and

400 color. The time between the population B peak in infections and the wastewater concentration  
401 peak is indicated by  $\Delta t$ . (c) The time between the population B peak in infections and the  
402 wastewater concentration peak (defined in (b)) when varying the fraction of cross-population  
403 contacts ( $\epsilon$ ) and the fraction of population B that is sampled ( $f_B$ ). Light blue dots indicate  
404 parameter regimes in which  $\Delta t$  is less than 1 generation time. In both (b) and (c), the simulation  
405 starts with a single infected individual in population A at time 0 and no infected individuals in  
406 population B. The parameters are set to population size  $N_A = N_B = 5000$ , basic reproduction  
407 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  
408 the probability of infection given contact  $\beta_A = \beta_B = \gamma_I R_0$ .

409  
410 We then weakened the assumption that population B had no contribution to the  
411 sampled wastewater and allowed a fraction of population B to be sampled by  
412 wastewater. For example, individuals in population B may commute to a workplace  
413 sampled by wastewater, or a fraction of population B may be connected at home. As the  
414 fraction of population B that is sampled by wastewater increases, the time between  
415 infection and wastewater peaks decreases; however, the fraction of B that is sampled  
416 has a much weaker effect on the wastewater peak time than the interaction parameter  
417 between the communities. For example, with the above parameters and equal  
418 population sizes, with 10% sampling of population B, the cross-population contacts  
419 must drop below 4%, and with 50% below 3% and with 100% below 2% before the  
420 wastewater concentration peak and the peak in the number of infections in population B  
421 drops below one generation time.

422

423 Varying the population size of both populations (with an equal population size in the two  
424 populations), the recovery rate or generation time, and the  $R_0$  of both populations (with  
425 an equal  $R_0$  in the two populations) has only a weak effect on the discrepancy between  
426 wastewater concentration and population B infection peaks (**Fig S 12**).

427  
428 When the unconnected population has a smaller size than the connected population,  
429 then even substantial interactions between the two populations can lead the wastewater  
430 concentration to peak before the infections in the unconnected population peaks (**Fig S**  
431 **12**). Both very weak and very strong interactions cause the wastewater peak to coincide  
432 with the peak number of infections in the unconnected population.

433  
434 Additionally, when the unconnected population has a smaller  $R_0$  than the connected  
435 population (which could occur for instance if the unconnected population was more  
436 spread out geographically and thus had a lower contact rate), the outbreak in population  
437 B peaks later than the wastewater peak leading to a discrepancy in the peak times (**Fig**  
438 **S 12**).

439

## 440 **Discussion**

441

### 442 **Equity of sewer connectivity**

443 Our analysis of available datasets revealed considerable variability in sewer connectivity  
444 within and between communities, across locations, demographics, and economic  
445 statuses in the United States. The western US had higher levels of sewer connectivity

446 than the eastern US. Across the US, lower connectivity to sewers was observed for  
447 American Indian and Alaska Native householders; White householders; non-Hispanic  
448 householders; older populations; and larger households in most census divisions. In the  
449 Pacific census division, lower connectivity to sewers was observed for Native Hawaiian  
450 and Other Pacific Islander householders specifically in areas inside of a metropolitan  
451 statistical area but outside of a central city. Across census divisions (with the exception  
452 of the Mountain census division), households with higher income were less connected  
453 to sewers, consistent with nation-wide observations [11]. The decrease in sewer  
454 connectivity by income could not be explained by collinearity between age and income  
455 as age and income exhibited a weak negative correlation (**Fig S 13**). These results are  
456 also consistent with the observation that in North Carolina, minorities and those with  
457 social vulnerabilities are overrepresented in the sewered population [21]. Aggregating  
458 by large metropolitan areas yielded similar trends as when aggregating by census  
459 divisions, except for lower connection for Hispanic householders in San Jose area and  
460 for households with lower income in San Francisco, Dallas, Chicago, and Tampa  
461 metropolitan areas, whereas these groups were better connected to sewers at the  
462 census division level. The differences in the MSA-level data suggested heterogeneity in  
463 connectivity across locations.

464

465 We have not attempted to determine why the observed differences in sewer connectivity  
466 exist. Many of the variables that we analyzed are likely to be correlated with one  
467 another.

468

## 469 **Geographic variation in sewer connectivity**

470 Large parts of Alaska and the Navajo Nation lacked plumbing. This is consistent with  
471 the finding from the Annual Report to Congress on Sanitation Deficiency Levels for  
472 Indian Homes and Communities in 2019 that approximately 20% of Indian homes in  
473 Alaska do not have sewage disposal and about 10% of Indian homes in the Navajo  
474 Nation do not have sewage disposal [40].

475  
476 State-level datasets revealed that sewer connectivity varied between and within  
477 communities. While in some states most areas had sewer connectivity (California, Iowa,  
478 Maryland, New Jersey, New York), in others there were large regions that lacked  
479 connectivity (Minnesota, Michigan, Florida). Asian and Hispanic populations were less  
480 connected to septic tanks across county subdivisions in Florida; however, there was  
481 considerable variability across the state and potential biases in the data. In contrast to  
482 the national data, in Minnesota, Florida, and California, counties with lower median  
483 household incomes were less connected to sewers, suggesting variability across the  
484 nation. This result in Florida is consistent with the observation that as income levels  
485 rose, household decentralized system usage declined in Florida [20]. However, we did  
486 not observe any significant correlation of income with septic usage in the Florida  
487 Department of Health onsite sewage treatment and disposal systems inspection  
488 permitting dataset, possibly due to not all septic tanks having a permit that is reported in  
489 this dataset.

490

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

495

## 496 **Data limitations**

497 Substantial data gaps and biases prevented a comprehensive analysis of disparities in  
498 sewer system connectivity within and between communities across the US, particularly  
499 for small communities, tribal lands and Alaska Native Villages, and state and local  
500 geographic scales. The current design of the American Housing Survey has too small of  
501 sample size to study state and local geographic scales [37]. Additionally, the AHS may  
502 have underestimated the fraction of households using septic systems due to its survey  
503 design (**Table 1**). The differences between Florida state data supports the need for  
504 detailed and reproducible longitudinal studies of sewer connectivity (both in developing  
505 sampling and analysis methods). Additionally, multiple datasets were over a decade old,  
506 and changes to which households are served by sewer and septic systems will have  
507 occurred since the datasets were collected.

508

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

514 population sizes served in New York state provides a template for how to create these  
515 sewer catchment maps from a combination of permitting, survey, and tax record data  
516 [16]. If possible, it would be helpful to consolidate and standardize existing data where it  
517 exists (for instance, sewage disposal permitting at the local or state levels). Additionally,  
518 if the EPA's request to expand the question on the ACS about access to plumbing  
519 facilities to additionally ask about the type of plumbing facility [20,41] is granted, then  
520 this would be a valuable dataset with the geographic and temporal resolution to study  
521 equity in sewer connectivity moving forward.

522

## 523 **Generalizability of wastewater-based epidemiology data in** 524 **light of heterogeneities in sewer connectivity**

525 Our modeling results suggest that even weak interactions between two communities  
526 allow wastewater monitoring in one community to serve as reliable proxy for the time of  
527 maximum infections in the other community when the population sizes and  $R_0$  of the two  
528 populations are comparable, but not when the unconnected population has a  
529 substantially lower population size or  $R_0$ . In the scenario with unequal population size,  
530 outbreaks that are seeded at the same time in the two populations in the absence of  
531 interactions peak earlier in the unconnected population due to its smaller size. With  
532 weak interactions, seeding of infections from population A to B occurs slower, leading to  
533 more coincidence in peaks; with strong interactions, then the dynamics of B are  
534 dominated by those in A, leading to more coincidence in peaks; at intermediate  
535 interaction strengths, infections are seeded earlier and peaks early, leading to the  
536 largest discrepancy in wastewater and infection peaks.



537

538 As the purpose of our model was to explore the impact of factors on the generalizability  
539 of wastewater rather than to accurately capture the dynamics in all scenarios, we chose  
540 the most simple compartmental model (SIR) and made simplifying assumptions,  
541 including that all infected individuals shed the same amount of pathogen genetic  
542 material into wastewater and shed only during the infected period; that all individuals  
543 contribute equally to wastewater; and that pathogen detection in wastewater has perfect  
544 sensitivity. We also note that compartmental models have previously been used to  
545 analyze wastewater data [42].

546

547 Our results suggest that in assessing the generalizability of wastewater data, it would be  
548 useful to estimate the extent of mobility between connected and unconnected  
549 communities. Interestingly, Ref [43] found no correlation between the size of a  
550 catchment area and the correlation of wastewater with case data for SARS-CoV-2,  
551 consistent with the result that interactions between connected and unconnected  
552 communities cause the disease dynamics to look similar to the wastewater data. In  
553 regions without sewer connectivity or with little interaction between connected and  
554 unconnected communities, wastewater data from neighboring communities will be less  
555 informative. In areas with low sewer connectivity in households, sampling wastewater  
556 outflow at frequently visited non-household locations (i.e., schools, offices, malls, etc.)  
557 may capture a more representative population.

558

559 **Additional considerations for equity in WBE**

560 In addition to inequities in sewer connectivity, the sewer locations used for wastewater  
561 sampling should be considered to promote demographic equity and ensuring the ability  
562 to capture spatiotemporal trends [21,44–46]. While we have focused on analysis on the  
563 US due to data availability, internationally disadvantaged populations are associated  
564 with lower access to sewers [47–49]. The ongoing development of wastewater sampling  
565 in non-sewered settings, for example in water channels in Las Vegas Valley [50], onsite  
566 sanitation facilities in Bangladesh [51], a refugee camp in Lebanon [52], and various  
567 non-sewered settings in low and middle income countries [53–55], represents a critical  
568 area for research and development.

569

## 570 **Conclusions**

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

577

## 578 **Acknowledgements**

579 We thank Cara Omana from the Minnesota Pollution Control Agency for generously  
580 sharing the Minnesota WINS data and Harry Campbell from the Utah Department of  
581 Environment Quality, Division of Water Quality for generously sharing the Utah MWPP

582 survey data and to both for answering questions about the datasets. We thank  
583 colleagues in the Grad lab and the Harvard T. H. Chan School of Public Health Center  
584 for Communicable Disease Dynamics for helpful discussions, particularly Stephen  
585 Kissler.

## 586 **References**

- 587 1. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A,  
588 et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection  
589 dynamics. *Nat Biotechnol.* 2020;38: 1164–1167.
- 590 2. Medema G, Been F, Heijnen L, Petterson S. Implementation of environmental  
591 surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities  
592 and challenges. *Curr Opin Environ Sci Health.* 2020;17: 49–71.
- 593 3. Pöyry T, Stenvik M, Hovi T. Viruses in sewage waters during and after a  
594 poliomyelitis outbreak and subsequent nationwide oral poliovirus vaccination  
595 campaign in Finland. *Appl Environ Microbiol.* 1988;54: 371–374.
- 596 4. Link-Gelles R, Lutterloh E, Schnabel Ruppert P, Backenson PB, St George K,  
597 Rosenberg ES, et al. Public Health Response to a Case of Paralytic Poliomyelitis in  
598 an Unvaccinated Person and Detection of Poliovirus in Wastewater - New York,  
599 June-August 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71: 1065–1068.
- 600 5. Kilaru P, Hill D, Anderson K, Collins MB, Green H, Kmush BL, et al. Wastewater  
601 Surveillance for Infectious Disease: A Systematic Review. *Am J Epidemiol.*  
602 2023;192: 305–322.
- 603 6. Boehm AB, Hughes B, Duong D, Chan-Herur V, Buchman A, Wolfe MK, et al.  
604 Wastewater concentrations of human influenza, metapneumovirus, parainfluenza,  
605 respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids  
606 during the COVID-19 pandemic: a surveillance study. *Lancet Microbe.* 2023.  
607 doi:10.1016/S2666-5247(22)00386-X

- 608 7. Baz-Lomba JA, Salvatore S, Gracia-Lor E, Bade R, Castiglioni S, Castrignanò E, et  
609 al. Comparison of pharmaceutical, illicit drug, alcohol, nicotine and caffeine levels in  
610 wastewater with sale, seizure and consumption data for 8 European cities. *BMC*  
611 *Public Health*. 2016;16: 1035.
- 612 8. Gushgari AJ, Venkatesan AK, Chen J, Steele JC, Halden RU. Long-term tracking of  
613 opioid consumption in two United States cities using wastewater-based  
614 epidemiology approach. *Water Res*. 2019;161: 171–180.
- 615 9. Holm RH, Pocock G, Severson MA, Huber VC, Smith T, McFadden LM. Using  
616 wastewater to overcome health disparities among rural residents. *Geoforum*.  
617 2023;144: 103816.
- 618 10. Analytics B. Wastewater for equitable Covid-19 monitoring. In: Medium [Internet]. 8  
619 Jul 2021 [cited 9 Feb 2023]. Available:  
620 [https://biobotanalytics.medium.com/wastewater-for-equitable-covid-19-monitoring-](https://biobotanalytics.medium.com/wastewater-for-equitable-covid-19-monitoring-e3b947d91e0a)  
621 [e3b947d91e0a](https://biobotanalytics.medium.com/wastewater-for-equitable-covid-19-monitoring-e3b947d91e0a)
- 622 11. Olesen SW, Young CCW, Duvallet C. White paper: The effect of septic systems on  
623 wastewater-based epidemiology. 2022. doi:10.5281/zenodo.7130726
- 624 12. U.S. Census Bureau. AHS 2021 National Public Use File. Available:  
625 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2021/AHS%202021%20National%20PUF%20v1.0%20CSV.zip)  
626 [surveys/ahs/2021/AHS%202021%20National%20PUF%20v1.0%20CSV.zip](https://www2.census.gov/programs-surveys/ahs/2021/AHS%202021%20National%20PUF%20v1.0%20CSV.zip)
- 627 13. US EPA Decentralized Wastewater Program}. *Septic Systems Fact Sheet*. 2008  
628 Oct.
- 629 14. Leker HG, MacDonald Gibson J. Relationship between race and community water  
630 and sewer service in North Carolina, USA. *PLoS One*. 2018;13: e0193225.

- 631 15. Meehan K, Jurjevich JR, Chun NMJW, Sherrill J. Geographies of insecure water  
632 access and the housing-water nexus in US cities. *Proc Natl Acad Sci U S A*.  
633 2020;117: 28700–28707.
- 634 16. Hill DT, Larsen DA. Using geographic information systems to link population  
635 estimates to wastewater surveillance data in New York State, USA. *PLOS Glob*  
636 *Public Health*. 2023;3: e0001062.
- 637 17. Rural Community Assistance Partnership. Still Living Without the Basics in the 21st  
638 Century: Analyzing the Availability of Water and Sanitation Services in the United  
639 States. 2004. Available: [http://opportunitylinkmt.org/wp-](http://opportunitylinkmt.org/wp-content/uploads/2015/07/Still-Living-Without-the-Basics-Water.pdf)  
640 [content/uploads/2015/07/Still-Living-Without-the-Basics-Water.pdf](http://opportunitylinkmt.org/wp-content/uploads/2015/07/Still-Living-Without-the-Basics-Water.pdf)
- 641 18. Hernandez A, Pierce G. The geography and socioeconomic characteristics of U.S.  
642 households reliant on private wells and septic systems. *J Am Water Resour Assoc*.  
643 2023;59: 1397–1412.
- 644 19. Brown J, Acey CS, Anthonj C, Barrington DJ, Beal CD, Capone D, et al. The effects  
645 of racism, social exclusion, and discrimination on achieving universal safe water  
646 and sanitation in high-income countries. *Lancet Glob Health*. 2023;11: e606–e614.
- 647 20. U.S. Environmental Protection Agency, Office of Water. Report to Congress on The  
648 Prevalence Throughout the U.S. of Low- and Moderate-Income Households Without  
649 Access to a Treatment Works and The Use by States of Assistance under Section  
650 603(c)(12) of the Federal Water Pollution Control Act. U.S. Environmental  
651 Protection Agency; 2021 Jul. Available:  
652 [https://www.epa.gov/system/files/documents/2022-01/low-mod-income-without-](https://www.epa.gov/system/files/documents/2022-01/low-mod-income-without-treatment_report-to-congress.pdf)  
653 [treatment\\_report-to-congress.pdf](https://www.epa.gov/system/files/documents/2022-01/low-mod-income-without-treatment_report-to-congress.pdf)

- 654 21. Hu XC, Reckling SK, Keshaviah A. Assessing health equity in wastewater  
655 monitoring programs: Differences in the demographics and social vulnerability of  
656 sewerred and unsewerred populations across North Carolina. medRxiv. 2023. p.  
657 2023.10.06.23296680. doi:10.1101/2023.10.06.23296680
- 658 22. U.S. Census Bureau. AHS 2019 National Public Use File. Available:  
659 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20National%20PUF%20v1.1%20CSV.zip)  
660 [surveys/ahs/2019/AHS%202019%20National%20PUF%20v1.1%20CSV.zip](https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20National%20PUF%20v1.1%20CSV.zip)
- 661 23. U.S. Census Bureau. AHS 2019 Metropolitan Public Use File. Available:  
662 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20Metropolitan%20PUF%20v1.0%20CSV.zip)  
663 [surveys/ahs/2019/AHS%202019%20Metropolitan%20PUF%20v1.0%20CSV.zip](https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20Metropolitan%20PUF%20v1.0%20CSV.zip)
- 664 24. U.S. Census Bureau, Department of Commerce Department of Housing and Urban  
665 Development. 2019 AHS Integrated National Sample: Sample Design, Weighting,  
666 and Error Estimation. U.S. Census Bureau, Department of Commerce Department  
667 of Housing and Urban Development; 2020 Sep. Available:  
668 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2019/2019%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)  
669 [surveys/ahs/2019/2019%20AHS%20National%20Sample%20Design,%20Weightin](https://www2.census.gov/programs-surveys/ahs/2019/2019%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)  
670 [g,%20and%20Error%20Estimation.pdf](https://www2.census.gov/programs-surveys/ahs/2019/2019%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)
- 671 25. U.S. Census Bureau, Department of Commerce Department of Housing and Urban  
672 Development. 2019 AHS Metropolitan Sample: Sample Design, Weighting, and  
673 Error Estimation. U.S. Census Bureau, Department of Commerce Department of  
674 Housing and Urban Development; 2020 Sep. Available:  
675 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2019/AHS%202019%20Metropolitan%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)

- 676 surveys/ahs/2019/2019%20AHS%20Metropolitan%20Sample%20Design,%20Weig  
677 hting,%20and%20Error%20Estimation.pdf
- 678 26. U.S. Census Bureau, Department of Commerce Department of Housing and Urban  
679 Development. 2021 AHS Integrated National Sample: Sample Design, Weighting,  
680 and Error Estimation. U.S. Census Bureau, Department of Commerce Department  
681 of Housing and Urban Development; 2022 Aug. Available:  
682 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2021/2021%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)  
683 [surveys/ahs/2021/2021%20AHS%20National%20Sample%20Design,%20Weightin](https://www2.census.gov/programs-surveys/ahs/2021/2021%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)  
684 [g,%20and%20Error%20Estimation.pdf](https://www2.census.gov/programs-surveys/ahs/2021/2021%20AHS%20National%20Sample%20Design,%20Weighting,%20and%20Error%20Estimation.pdf)
- 685 27. US Census Bureau. American Community Survey Information Guide. [cited 5 Jan  
686 2023]. Available: [https://www.census.gov/programs-surveys/acs/library/information-](https://www.census.gov/programs-surveys/acs/library/information-guide.html)  
687 [guide.html](https://www.census.gov/programs-surveys/acs/library/information-guide.html)
- 688 28. General Facts and Statistics about Wastewater in Florida. [cited 10 Feb 2023].  
689 Available: [https://floridadep.gov/water/domestic-wastewater/content/general-facts-](https://floridadep.gov/water/domestic-wastewater/content/general-facts-and-statistics-about-wastewater-florida)  
690 [and-statistics-about-wastewater-florida](https://floridadep.gov/water/domestic-wastewater/content/general-facts-and-statistics-about-wastewater-florida)
- 691 29. Florida Department of Health. Onsite Sewage Locations in Florida - 2012. Available:  
692 [https://fgdl.org/zips/metadata/xml/septic\\_jun12.xml](https://fgdl.org/zips/metadata/xml/septic_jun12.xml)
- 693 30. Environmental Protection Agency. Clean Watersheds Needs Survey Detailed  
694 Scope and Methods. Environmental Protection Agency; 2012. Available:  
695 [https://www.epa.gov/sites/default/files/2015-](https://www.epa.gov/sites/default/files/2015-10/documents/cwns_2012_detailed_scope_and_methods-508.pdf)  
696 [10/documents/cwns\\_2012\\_detailed\\_scope\\_and\\_methods-508.pdf](https://www.epa.gov/sites/default/files/2015-10/documents/cwns_2012_detailed_scope_and_methods-508.pdf)
- 697 31. Environmental Protection Agency. Clean Watersheds Needs Survey 2012 Report to  
698 Congress. Environmental Protection Agency; 2016. Available:



- 699 <https://www.epa.gov/sites/default/files/2015->  
700 [12/documents/cwns\\_2012\\_report\\_to\\_congress-508-opt.pdf](https://www.epa.gov/sites/default/files/2015-12/documents/cwns_2012_report_to_congress-508-opt.pdf)
- 701 32. U.S. Census Bureau. Guidance on Using 2020 Island Areas Censuses Data. 20 Oct  
702 2022 [cited 31 Mar 2023]. Available: [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/2020-iac-guidance.pdf)  
703 [surveys/decennial/2020/technical-documentation/island-areas-tech-docs/2020-iac-](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/2020-iac-guidance.pdf)  
704 [guidance.pdf](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/2020-iac-guidance.pdf)
- 705 33. U.S. Census Bureau. 2020 Island Areas Censuses (IAC) Demographic Profile  
706 Summary File. U.S. Census Bureau; 2022 Oct. Available:  
707 [https://www2.census.gov/programs-surveys/decennial/2020/technical-](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/demographic_profile/2020-iac-dpsf-technical-documentation.pdf)  
708 [documentation/island-areas-tech-docs/demographic\\_profile/2020-iac-dpsf-](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/demographic_profile/2020-iac-dpsf-technical-documentation.pdf)  
709 [technical-documentation.pdf](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/island-areas-tech-docs/demographic_profile/2020-iac-dpsf-technical-documentation.pdf)
- 710 34. U.S. Department of Housing and Urban Development and U.S. Census Bureau.  
711 2013 AHS Definitions. In: 2013 AHS Definitions [Internet]. [cited 9 Feb 2023].  
712 Available: [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2013/2013%20AHS%20Definitions.pdf)  
713 [surveys/ahs/2013/2013%20AHS%20Definitions.pdf](https://www2.census.gov/programs-surveys/ahs/2013/2013%20AHS%20Definitions.pdf)
- 714 35. TIGER/Line® Shapefiles. [cited 25 Apr 2023]. Available:  
715 <https://www.census.gov/cgi-bin/geo/shapefiles/index.php>
- 716 36. U.S. Census Bureau. Metropolitan and Micropolitan About. In: Metropolitan and  
717 Micropolitan About [Internet]. 22 Nov 2021 [cited 9 Feb 2023]. Available:  
718 <https://www.census.gov/programs-surveys/metro-micro/about.html>
- 719 37. Schwartz MB. 2007 American Community Survey: A Comparison to Selected  
720 Housing and Financial Characteristics for the United States from the 2007 American  
721 Housing Survey. Housing and Household Economic Statistics Division U.S. Census

- 722 Bureau; Available: <https://www.huduser.gov/portal/datasets/ahs/acs->  
723 [ahscomparisonreport.pdf](https://www.huduser.gov/portal/datasets/ahs/ahscomparisonreport.pdf)
- 724 38. U.S. Census Bureau. Poverty Thresholds. 12 Sep 2023 [cited 7 Dec 2023].  
725 Available: [https://www.census.gov/data/tables/time-series/demo/income-](https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html)  
726 [poverty/historical-poverty-thresholds.html](https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html)
- 727 39. U.S. Census Bureau. AHS 2013 National Public Use File. Available:  
728 [https://www2.census.gov/programs-](https://www2.census.gov/programs-surveys/ahs/2013/AHS%202013%20National%20PUF%20v2.0%20CSV.zip)  
729 [surveys/ahs/2013/AHS%202013%20National%20PUF%20v2.0%20CSV.zip](https://www2.census.gov/programs-surveys/ahs/2013/AHS%202013%20National%20PUF%20v2.0%20CSV.zip)
- 730 40. Indian Health Service; Office of Environmental Health and Engineering; Division of  
731 Sanitation Facilities Construction. Annual Report To the Congress of the United  
732 States On Sanitation Deficiency Levels for Indian Homes and Communities: Fiscal  
733 Year 2019. Rockville (MD): Indian Health Service (US); 2020.
- 734 41. Environmental Protection Agency. EPA's Proposal: Justification for Requested  
735 Revision, Addition or Deletion to Questions on the American Community Survey  
736 (ACS). 2018. Available:  
737 <https://foiaonline.gov/foiaonline/action/public/submissionDetails?trackingNumber=E>  
738 [PA-HQ-2018-010667&type=request](https://foiaonline.gov/foiaonline/action/public/submissionDetails?trackingNumber=E)
- 739 42. McMahan CS, Self S, Rennert L, Kalbaugh C, Kriebel D, Graves D, et al. COVID-19  
740 wastewater epidemiology: a model to estimate infected populations. *Lancet Planet*  
741 *Health*. 2021;5: e874–e881.
- 742 43. Duvallet C, Wu F, McElroy KA, Imakaev M, Endo N, Xiao A, et al. Nationwide  
743 Trends in COVID-19 Cases and SARS-CoV-2 RNA Wastewater Concentrations in  
744 the United States. *ACS EST Water*. 2022. doi:10.1021/acsestwater.1c00434

- 745 44. Medina CY, Kadonsky KF, Roman FA Jr, Tariqi AQ, Sinclair RG, D'Aoust PM, et al.  
746 The need of an environmental justice approach for wastewater based epidemiology  
747 for rural and disadvantaged communities: A review in California. *Curr Opin Environ*  
748 *Sci Health*. 2022;27: 100348.
- 749 45. Haak L, Delic B, Li L, Guarin T, Mazurowski L, Dastjerdi NG, et al. Spatial and  
750 temporal variability and data bias in wastewater surveillance of SARS-CoV-2 in a  
751 sewer system. *Sci Total Environ*. 2022;805: 150390.
- 752 46. Naughton CC, Roman FA Jr, Alvarado AGF, Tariqi AQ, Deeming MA, Bibby K, et  
753 al. Show us the Data: Global COVID-19 Wastewater Monitoring Efforts, Equity, and  
754 Gaps. *bioRxiv. medRxiv*; 2021. doi:10.1101/2021.03.14.21253564
- 755 47. Adhikari S, Halden RU. Opportunities and limits of wastewater-based epidemiology  
756 for tracking global health and attainment of UN sustainable development goals.  
757 *Environ Int*. 2022;163: 107217.
- 758 48. Wolf J, Johnston RB, Ambelu A, Arnold BF, Bain R, Brauer M, et al. Burden of  
759 disease attributable to unsafe drinking water, sanitation, and hygiene in domestic  
760 settings: a global analysis for selected adverse health outcomes. *Lancet*. 2023;401:  
761 2060–2071.
- 762 49. WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and  
763 Hygiene. Progress on Household Drinking Water, Sanitation and Hygiene. WHO  
764 and UNICEF; 2021.
- 765 50. Gerrity D, Papp K, Dickenson E, Ejjada M, Marti E, Quinones O, et al.  
766 Characterizing the chemical and microbial fingerprint of unsheltered homelessness  
767 in an urban watershed. *Sci Total Environ*. 2022;840: 156714.

- 768 51. Jakariya M, Ahmed F, Islam MA, Al Marzan A, Hasan MN, Hossain M, et al.  
769 Wastewater-based epidemiological surveillance to monitor the prevalence of SARS-  
770 CoV-2 in developing countries with onsite sanitation facilities. *Environ Pollut*.  
771 2022;311: 119679.
- 772 52. Tokajian S, Moghnieh R, Salloum T, Arabaghian H, Alousi S, Moussa J, et al.  
773 Extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* in wastewaters and  
774 refugee camp in Lebanon. *Future Microbiol*. 2018;13: 81–95.
- 775 53. Gwenzi W. Wastewater, waste, and water-based epidemiology (WWW-BE): A novel  
776 hypothesis and decision-support tool to unravel COVID-19 in low-income settings?  
777 *Sci Total Environ*. 2022;806: 150680.
- 778 54. Gina Pocock, Leanne Coetzee, Janet Mans, Maureen Taylor, and Bettina Genthe.  
779 Proof of Concept Study Application of wastewater-based surveillance to monitor  
780 SARS-CoV-2 prevalence in South African communities. *Water Research*  
781 Commission Report. 2020. Available: [https://wrcwebsite.azurewebsites.net/wp-](https://wrcwebsite.azurewebsites.net/wp-content/uploads/mdocs/TT%20832-20%20final%20web.pdf)  
782 [content/uploads/mdocs/TT%20832-20%20final%20web.pdf](https://wrcwebsite.azurewebsites.net/wp-content/uploads/mdocs/TT%20832-20%20final%20web.pdf)
- 783 55. Barnes KG, Levy JI, Gauld J, Rigby J, Kanjerwa O, Uzzell CB, et al. Utilizing river  
784 and wastewater as a SARS-CoV-2 surveillance tool in settings with limited formal  
785 sewage systems. *Nat Commun*. 2023;14: 7883.
- 786 56. U.S. Department of Housing and Urban Development and U.S. Census Bureau.  
787 Appendix B: Sample Design and Weighting: 2013. U.S. Department of Housing and  
788 Urban Development and U.S. Census Bureau; 2013. Available:  
789 <https://www2.census.gov/programs->

- 790 surveys/ahs/2013/2013%20AHS%20National%20Sample%20Design%20and%20  
791 Weighting.pdf
- 792 57. Onsite Sewage FAQ - Permitting. [cited 10 Feb 2023]. Available:  
793 <https://floridadep.gov/water/onsite-sewage/content/onsite-sewage-faq-permitting>
- 794 58. Florida Department of Health. Septic System Owner's Guide. In: floridahealth.gov  
795 [Internet]. [cited 10 Feb 2023]. Available:  
796 [https://www.floridahealth.gov/environmental-health/onsite-](https://www.floridahealth.gov/environmental-health/onsite-sewage/_documents/septic-folder-8x11-link2.pdf)  
797 [sewage/\\_documents/septic-folder-8x11-link2.pdf](https://www.floridahealth.gov/environmental-health/onsite-sewage/_documents/septic-folder-8x11-link2.pdf)
- 798 59. Utah Department of Environmental Quality Water Quality. Municipal Wastewater  
799 Planning Program (MWPP). In: Municipal Wastewater Planning Program (MWPP)  
800 [Internet]. 3 Feb 2023 [cited 10 Feb 2023]. Available: [https://deq.utah.gov/water-](https://deq.utah.gov/water-quality/municipal-wastewater-planning-program-mwpp)  
801 [quality/municipal-wastewater-planning-program-mwpp](https://deq.utah.gov/water-quality/municipal-wastewater-planning-program-mwpp)
- 802 60. Campbell H. 2021 MWPP Survey. In: 2021 MWPP Survey [Internet]. 27 Oct 2021  
803 [cited 10 Feb 2023]. Available: [https://documents.deq.utah.gov/water-](https://documents.deq.utah.gov/water-quality/engineering/municipal-wastewater-planning-program/DWQ-2021-026722.pdf)  
804 [quality/engineering/municipal-wastewater-planning-program/DWQ-2021-026722.pdf](https://documents.deq.utah.gov/water-quality/engineering/municipal-wastewater-planning-program/DWQ-2021-026722.pdf)
- 805 61. Minnesota Pollution Control Agency. 2021 Wastewater Infrastructure Needs  
806 Instructions and Frequently Asked Questions (FAQ). Apr 2021 [cited 31 Mar 2023].  
807 Available: <https://www.pca.state.mn.us/sites/default/files/wq-wwtp3-06.pdf>
- 808 62. Robinson C. 2017 SSTS Annual Report Subsurface Sewage Treatment Systems in  
809 Minnesota. Minnesota Pollution Control Agency; 2018 Apr. Available:  
810 <https://www.pca.state.mn.us/sites/default/files/wq-wwists1-58.pdf>
- 811 63. U.S. Environmental Protection Agency. 2012 Clean Watersheds Needs Survey.  
812 Available: <https://ordspub.epa.gov/ords/cwns2012/f?p=cwns2012:25>

- 813 64. Weidhaas J, Aanderud ZT, Roper DK, VanDerslice J, Gaddis EB, Ostermiller J, et  
814 al. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden  
815 in sewersheds. *Sci Total Environ.* 2021;775: 145790.
- 816 65. U.s E. WaterSense Statistics and Facts. 11 May 2022 [cited 3 Apr 2023]. Available:  
817 <https://www.epa.gov/watersense/statistics-and-facts>  
818

## 819 **Supporting information**

### 820 **S1 Appendix. Supplementary method**

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

### 824 **S1 Appendix. Supplementary methods.**

#### 825 **Dataset descriptions and data cleaning**

##### 826 **U.S. Census Bureau American Housing Survey 2019 and 2021**

827 The American Housing Survey (AHS) National and Metropolitan Public Use Files were  
828 downloaded from Ref. [12,22,23]. The sample design, weighting, and error estimation  
829 survey is described in Refs. [24–26]. Briefly, the AHS is a biannual survey that assesses  
830 housing characteristics. Two samples of households are chosen every survey cycle, the  
831 Integrated National Sample and the Independent Metropolitan Sample. The Integrated  
832 National Sample consists of nationally representative households and includes an  
833 oversample of each of 15 largest metropolitan areas. The Independent Metropolitan  
834 Sample consists of representative households in 20 large metropolitan areas that are  
835 representative of the next largest 50 metropolitan areas. Ten of the 20 large  
836 metropolitan areas are sampled every other survey cycle (10 metropolitan areas  
837 surveyed every 2 years). The Integrated National Sample in 2019 included 86,257  
838 selected representative housing units, of which about 63,500 met the AHS definition of  
839 a housing unit and were able to be interviewed, and in 2021 included 95,295 selected  
840 representative housing units, of which about 64,100 were interviewed. The Integrated

841 Metropolitan Samples in 2019 and 2021 included around 3000 selected representative  
842 housing units in each metropolitan area, of which around 1000-3000 were eligible and  
843 were able to be interviewed (same criteria as in the Integrated National Sample).

844  
845 Geographic data of each household available in the Public Use File included the census  
846 division; core based statistical area (CBSA) for the 15 largest metropolitan areas and  
847 the 20 large metropolitan areas in the metropolitan data. All other households are  
848 classified as in “all other metropolitan areas” or “not in a metropolitan area” (all other  
849 locations).

850  
851 The race, ethnicity, and age of the householder (the person or one of the persons who  
852 is an owner or renter of the unit) was used to categorize the race, ethnicity, and age of  
853 the household. Due to the low numbers of households where the householder  
854 responded as having more than one race, we focused on households where  
855 householders identified as having a single race.

856  
857 **U.S. Census Bureau American Housing Survey National Sample 2013**

858 The American Housing Survey National (2013) Use File was downloaded from Ref. [39].  
859 The sample design, weighting, and error estimation survey is described in Ref [56]. The  
860 National Sample included 84,400 selected housing units representative nationally and  
861 including a supplemental sample of 15,553 housing units in the Chicago, Detroit, New  
862 York City, Northern New Jersey, and Philadelphia metropolitan statistical areas, of  
863 which about 71,600 met the definition of a housing unit and were able to be interviewed.



864

865 Geographic data of each household available in the Public Use File included the census  
866 division; census region; central city, urban, or rural status; and select metropolitan area  
867 codes. All other housing units are classified into “not in a metropolitan area” or had their  
868 code suppressed for confidentiality [34].

869

## 870 **Florida Department of Health onsite sewage treatment and disposal** 871 **system inspections**

872 The locations of onsite sewage treatment and disposal systems inspected by Florida  
873 Department of Health reported June 2012 was downloaded from Ref. [29]. As of 2022,  
874 an inspection permit is required for construction [57] and inspection is recommended  
875 every 3 to 5 years after [58]. The number of systems inspected that are active at the  
876 latest inspection (after removing repeat inspections) is 601,430. The years of inspection  
877 are primarily 1998-2011, with some outliers as early as 1900, which may be old septic  
878 systems or typos, and as late as 2200, which are likely typos.

879

## 880 **Utah Municipal Wastewater Planning Program (MWPP) Survey 2021**

881 Data from the Utah Municipal Wastewater Planning Survey was provided by the Utah  
882 Department of Environment Quality, Division of Water Quality. The dataset includes an  
883 estimate of the population receiving collection by each utility that owns or operates a  
884 sanitary sewerage system and its latitude and longitude. The data gathered from  
885 treatment plant operators and participation is reported as mandatory [59], although only  
886 71% of contacts responded [60]. We filtered the data to include any facility that has

887 collection (this included collection only, treatment and collection, or small lagoons which  
888 are often collection and treatment). Manual correction of facility longitude and latitude is  
889 described in the supplementary methods.

890  
891 Manual editing of facility longitude and latitude was done for 5 facilities' that did not fall  
892 within the state of Utah when combining with the 2021 TIGER shapefile (BRIAN HEAD  
893 TOWN, CANYONG LAND IMPROVEMENT DISTRICT, HILL AFB (AMERICAN  
894 WATER), JORDANELLE SSD, PLAIN CITY) and 1 facility that did not report a latitude  
895 and longitude (PANGUITCH LAKE S. S. D.). For these facilities, the latitude and  
896 longitude were manually edited to be that of the location found in a Google Maps search  
897 of the facility name. If the Google Maps search found a city, then an arbitrarily chosen  
898 location in that city was used to obtain a latitude and longitude.

899  
900 **Minnesota Wastewater Infrastructure Needs Survey (WINS) 2021**

901 Data for the Minnesota Wastewater Infrastructure Needs Survey [61] was provided by  
902 the Minnesota Pollution Control Agency. Communities (places and townships)  
903 responded to the question "Does your community have a collection system?" with "yes",  
904 "no", or no response. The survey did not receive a response from every community in  
905 Minnesota.

906  
907 The Census Designated Place (CDP) name was obtained from the reported community  
908 name with manual changes reported in **Table S 5**.

909

## 910 **Minnesota Subsurface Sewage Treatment Systems 2017**

911 The Minnesota Subsurface Sewage Treatment Systems (SSTS) dataset reported the  
912 number of SSTS by local government units with a known SSTS program. Data was  
913 taken from Ref. [62].

914 Two hundred and eleven out of 2018 known SSTS programs submitted an annual  
915 report. Not all SSTS programs were able to be identified and contacts were not always  
916 provided. Fourteen reported having zero SSTS within their jurisdiction despite permitting  
917 SSTS in 2017. Data were reported to sometimes be estimates.

918

## 919 **U.S. Environmental Protection Agency Clean Watersheds Needs**

### 920 **Survey 2012**

921 Data for the most recent Environmental Protection Agency (EPA) Clean Watersheds Needs  
922 Survey (CWNS) survey (2012) [31] was downloaded from [63] and extracted from a Microsoft  
923 Access database file to csv files using the mdbtools package  
924 (<https://github.com/mdbtools/mdbtools>). The survey is administered by the EPA and the States  
925 to assess funding needs for unfunded capital costs of treatment works projects that (1) address  
926 a water quality or related public health problem existing as of January 1, 2012, or is expected to  
927 occur within the next 20 years and (2) meet the CWNS documentation criteria. CWNS  
928 documentation criteria includes (1) description and location of the problem, (2) site-specific  
929 solution to the problem, and (3) detailed cost information for implementing the solution. Data is  
930 collected about (1) publicly owned wastewater collection and treatment facilities, (2) combined  
931 sewer overflow control facilities, (3) stormwater management activities, and (4) decentralized  
932 wastewater treatment facilities. Data is collected on (a) estimated needs, costs and technical  
933 information, (b) facility location and contact information, and (c) facility population served, flow,

934 effluent, and unit process information. Supporting documentation was required for the entered  
935 data with some exceptions (for instance small communities could use a more simple form), and  
936 quality control was performed by reviewers to compare documentation with data entered in the  
937 system to ensure consistency of technical and cost data [30]. Data entry occurred from January  
938 to December 2012.

939  
940 There was large variability in the level of effort and resources that each state put into the survey.  
941 New York, California, Florida, New Jersey, Maryland, Iowa, Minnesota, and Michigan likely  
942 invested heavily in participating in the survey or already had similar state-level data collection  
943 systems in place for getting comprehensive responses [29,61,62].

944  
945 Facilities not included in the dataset are those in South Carolina, the Northern Mariana Islands,  
946 and American Samoa as they did not participate in the survey; facilities whose projects did not  
947 have documented solutions or cost estimates; privately owned wastewater facilities that serve  
948 privately owned industrial facilities, military installations, national parks, or other federal facilities;  
949 facilities on tribal lands and in Alaskan Native Villages, which were separately surveyed by the  
950 Indian Health Service; and facilities with projects that had existing funding prior to January 1,  
951 2012. Additionally, larger facilities were more likely to be captured by the survey due to the  
952 resource requirements for obtaining the right documentation. To address this bias, the survey  
953 had decreased documentation needs for small communities (population  $\leq 10,000$ ); despite this,  
954 the small community facilities were still underrepresented (pg. 2 of Ref. [31]).

955  
956 We filtered the facilities to the current facilities of type “collection: combined sewers” or  
957 “collection: separate sewers”. The population receiving collection is taken from the reported total  
958 number of residents who are connected to a sewer system which empties into a treatment plant.  
959 This does not include non-resident populations, populations served by acceptable decentralized

960 wastewater treatment systems, or populations connected to sewers that do not discharge to a  
961 treatment plant.

962

### 963 **U.S. Census Bureau Island Areas Decennial Survey 2020**

964 The Island Areas Decennial Survey surveys housing, social, and economic information  
965 for all housing units in Guam, the Northern Mariana Islands, the Virgin Islands, and  
966 American Samoa [33]. Data of general demographic characteristics (DP01), selected  
967 economic characteristics (DP03), and selected housing characteristics (DP04) were  
968 downloaded from <https://data.census.gov/>. Housing characteristics included data for  
969 both occupied and vacant housing units.

970

### 971 **U.S. Census Bureau American Community Survey**

972 The American Community Survey (ACS) is a monthly survey of a representative  
973 subsample of addresses in the US [27] on characteristics of populations and  
974 households. The ACS 5 year estimates (average over 5 years, less subject to  
975 fluctuations) for selected social characteristics (DP02), selected economic  
976 characteristics (DP03), and demographic and housing estimates (DP05) stratified  
977 county or county subdivision for select years were downloaded from  
978 <https://data.census.gov/>.

979

## 980 **Data analysis and visualization**

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

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

990

### 991 **Utah Municipal Wastewater Planning Program Survey analysis**

992 The population in each county receiving collection was calculated by summing the population  
993 receiving collection across all reported facilities in that county. The population in each county  
994 subdivision receiving collection was calculated by summing the population receiving collection  
995 across all reported facilities whose latitude and longitude fell within that county subdivision's  
996 boundaries from the 2021 TIGER shapefile. To calculate the fraction of a county or county  
997 subdivision receiving sewer collection, the population receiving collection was divided by the  
998 population size of each county or county subdivision reported by the 2021 American Community  
999 Survey 5-year estimate. Only county subdivisions with at least 5 households and 20 population  
1000 size were used.

1001

### 1002 **U.S. Environmental Protection Agency Clean Watersheds Needs**

#### 1003 **Survey analysis**

1004 To compare the EPA CWNS and AHS data in Core Based Statistical Areas (CBSA), counties  
1005 were categorized into CBSAs using the February 2013 U.S. Census Bureau file. To calculate  
1006 the fraction of a county receiving collection, the population size receiving collection in that  
1007 county was divided by the population size of each county reported by the 2012 American

1008 Community Survey 5-year estimate. We flagged counties that had no facilities reported, which  
1009 could either be because there were no facilities in that county or because of a lack of reporting,  
1010 and counties that had >100% sewer connectivity by population (potentially due to misreporting  
1011 or collection occurring across multiple counties) were set to 100% connectivity.

1012

## 1013 **U.S. Census Bureau Island Areas Decennial Survey analysis**

1014 To calculate the fraction of a Census Designated Place (CDP) connected to sewers, the number  
1015 of households connected to sewers was divided by the total number of households in a CDP.

1016 Only CDPs with at least 5 households and 20 population size were used.

1017

## 1018 **Generation of maps**

1019 TIGER shapefiles from the U.S. Census Bureau in the closest matching year to the dataset  
1020 were used for generating maps. Maps were generated using the geopandas package (version  
1021 0.12.2) in Python (version 3.11.0) and were displayed in the Mercator projection.

1022

## 1023 **Simulations of interacting populations and wastewater**

### 1024 **sampling**

1025 Two interacting populations are modeled in a compartmental model with susceptible, infected,  
1026 and recovered (SIR) compartments for each population. The number of susceptible, infected,  
1027 and recovered individuals in population A ( $S_A$ ,  $I_A$ , and  $R_A$ , respectively) and in population B ( $S_B$ ,  
1028  $I_B$ , and  $R_B$ , respectively) are described by:

1029

$$1030 \quad \frac{dS_A}{dt} = -\beta_A(1 - \varepsilon)I_A \frac{S_A}{N_A} - \beta_B \varepsilon I_B \frac{S_A}{N_A}$$

1031 
$$\frac{dS_B}{dt} = -\beta_B(1 - \varepsilon)I_B \frac{S_B}{N_B} - \beta_A \varepsilon I_A \frac{S_B}{N_B}$$

1032 
$$\frac{dI_A}{dt} = \beta_A(1 - \varepsilon)I_A \frac{S_A}{N_A} + \beta_B \varepsilon I_B \frac{S_A}{N_A} - \gamma_I I_A$$

1033 
$$\frac{dI_B}{dt} = \beta_B(1 - \varepsilon)I_B \frac{S_B}{N_B} + \beta_A \varepsilon I_A \frac{S_B}{N_B} - \gamma_I I_B$$

1034 
$$\frac{dR_A}{dt} = \gamma_I I_A$$

1035 
$$\frac{dR_B}{dt} = \gamma_I I_B$$

1036

1037 where  $N_A$  and  $N_B$  are the number of individuals in populations A and B respectively,  $\beta_A$  and  $\beta_B$

1038 are the overall contact rates of populations A and B respectively times the probability of infection

1039 given contact,  $\gamma_I$  is the rate of recovery.  $\varepsilon$  is the fraction of the total contacts made by individuals

1040 in a population that occurs with individuals in the other population and it describes the

1041 interaction strength between the two populations.  $\varepsilon$  can take on values between 0 and 1, where

1042  $\varepsilon=0$  indicates that all interactions occur within each population and none between populations,

1043 and  $\varepsilon=1$  indicates that all interactions occur between the populations and none within

1044 populations. The base parameters are set at  $N_A=N_B=5000$ ,  $\gamma_I=0.18$  inverse days,  $R_0^A = R_0^B =$

1045  $\frac{\beta_A}{\gamma_I} = \frac{\beta_B}{\gamma_I} = 1.5$ , which represent an outbreak similar to COVID-19, but the parameters are varied

1046 in sensitivity analyses. The initial condition was a single infected individual in population A ( $I_A=1$ )

1047 and all other individuals were susceptible ( $S_A=N-1$ ,  $S_B=N$ ,  $I_B=0$ ,  $R_A=0$ ,  $R_B=0$ ). The differential

1048 equations are solved deterministically using the `integrate.odeint` function in the `scipy` package

1049 (version 1.6.2) in Python using a timestep of 0.1 days.

1050

1051 The concentration  $C$  of the pathogen in wastewater is given by

1052



1053

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

1054

1055 where  $n_A$  and  $n_B$  are the number of copies of pathogen genetic material (DNA or RNA) shed per  
1056 day by population A and B, respectively, into the sampled wastewater and  $v_A$  and  $v_B$  are the  
1057 volume of wastewater produced per day by population A and B, respectively, into the sampled  
1058 wastewater.

1059

1060 The number of copies of pathogen genetic material,  $n_A$  and  $n_B$ , shed per day by population A  
1061 and B, respectively, into the sampled wastewater is given by

1062

1063

$$n_A = nI_A f_A$$

1064

$$n_B = nI_B f_B$$

1065

1066 Where  $n$  is the number of copies of pathogen genetic material shed per infected person per day  
1067 (assumed to be the same for all infected individuals) and  $f_A$  and  $f_B$  are the fraction of total shed  
1068 pathogen genetic material by population A and B respectively that is sampled by wastewater  
1069 (determined by the sewer connectivity of each of the populations).

1070

1071 The volume of wastewater,  $v_A$  and  $v_B$ , produced per day by population A and B, respectively,  
1072 into the sampled wastewater is given by

1073

$$v_A = vN_A f_A$$

1074

$$v_B = vN_B f_B$$

1075

1076 where  $v$  is the volume of wastewater produced per person per day (assumed to be the same for  
1077 all individuals). The value used for  $n$  was  $5 \times 10^{6.7}$  gene copies/person/day (estimate taken from

1078 Ref. [64] by multiplying  $10^{4.7}$  gene copies/mL feces by 500 mL feces/day/person) and for  $v$  was  
1079 20 liters wastewater/person/day (per person daily wastewater usage approximated from per  
1080 household daily wastewater usage from Ref. [65] assuming 4 people per household).

1081  
1082 In addition to the assumptions mentioned above, the model makes the following additional  
1083 assumptions: The probability of an infected individual shedding pathogen genetic material to  
1084 wastewater is 1. Given that there is pathogen genetic material present in wastewater, the  
1085 probability of detecting it is 1. Shedding of pathogen genetic material occurs only when an  
1086 individual is infected. There are no stormwater contributions to wastewater to dilute the  
1087 pathogen concentration.

1088  
1089 To calculate the time of the maximum number of infections in a population and the time of the  
1090 maximum concentration of pathogen gene content in wastewater data, we numerically  
1091 calculated the time derivative at the midpoints of the time points as the difference between  
1092 consecutive datapoints divided by the time interval and took the first time point at which the  
1093 derivative crossed 0.

## 1094 **S2 Appendix. Supplementary analyses.**

### 1095 **Compilation of data and assessment of data completeness**

1096 As the AHS does not provide finer scale geographic information than the census  
1097 division for households outside of the 35 selected large metropolitan areas, we used the  
1098 American Community Survey, which reports the percentage of occupied housing units  
1099 lacking complete plumbing facilities by county and which allowed us to define the lower  
1100 bound on the percentage of the population not connected to sewers. Additionally, we  
1101 identified several state-level datasets, including locations of permitted septic tanks in

1102 Florida, sewer locations and populations served in Utah and Minnesota, and the 2012  
1103 EPA Clean Watersheds Needs Survey of sewer locations and populations served for 49  
1104 states (South Carolina did not participate). However, many of these state-level datasets  
1105 are incomplete or potentially biased and thus only allow qualitative analyses. Finally, the  
1106 2020 US Census' Island Areas Decennial survey provided information the percentage of  
1107 households connected to sewer by Census Designated Place in the island areas of  
1108 American Samoa, Guam, the Northern Mariana Islands, and the Virgin Islands (see  
1109 **Error! Reference source not found.** and **Table S 4** for a summary of datasets).

1110  
1111 We checked the consistency of the Florida Department of Health onsite sewage  
1112 treatment and disposal systems inspections data reported in 2012 with other sources.  
1113 We noted that an estimated 7% of households across the state were connected to  
1114 septic tanks, but that this was substantially less than the Florida Department of  
1115 Environmental Protection's estimate that approximately one third of Florida's population  
1116 uses septic tanks [28] even after accounting for potential correlations of septic usage  
1117 with household size (approximate increase in household size of 10% for a 50% increase  
1118 in septic connectivity (**Fig S 7**) cannot explain the almost 5 time difference in septic  
1119 connectivity across the two sources), suggesting that more households may be on  
1120 septic tanks, and thus less on sewer systems. This discrepancy may exist because not  
1121 every septic tank had a permit recorded in the state system, because there were  
1122 changes since 2012, or because of uncertainty in the Florida Department of  
1123 Environmental Protection's estimate. Thus, we focused our analysis on relative  
1124 comparisons across county subdivisions (Fig 4).

1125  
1126 The 2012 EPA Clean Watersheds Needs Survey dataset includes voluntary  
1127 submissions of the locations of publicly owned wastewater collection and treatment  
1128 facilities and their estimated population served. The following states had among the  
1129 highest reported needs and survey participation level (pg. 6 and Table A-1 of Ref. [31])  
1130 and were the focus of our analyses: New York, California, Florida, New Jersey,  
1131 Maryland, Iowa, Minnesota, and Michigan. Comparison of Florida in the EPA and  
1132 Florida Department of Public Health datasets showed consistent spatial trends (less  
1133 connection to sewers and more connection to septic systems in the panhandle and  
1134 inland) (Fig 4 and Fig S 11). Additionally, in both datasets, counties with more  
1135 connection to sewers are more Asian (EPA: Pearson  $r = 0.53$  ( $p = 4.2 \times 10^{-4}$ ), Florida  
1136 Department of Public Health: Pearson  $r = 0.31$  ( $p = 3.2 \times 10^{-7}$ )). Comparison of  
1137 Minnesota in the EPA dataset, Minnesota Wastewater Infrastructure Needs Survey, and  
1138 Minnesota SSTS datasets also showed consistent spatial trends (more connection to  
1139 sewers and less connection to subsurface sewage treatment systems around  
1140 Minneapolis) (Fig S 14). We found a significant correlation between the AHS and EPA  
1141 datasets for the largest 35 metropolitan areas in the states with good data collection  
1142 (Fig S 15), despite deviations from a one-to-one correspondence suggesting overall  
1143 systematic biases in the data.

1144  
1145 The 2021 Utah Municipal Wastewater Planning Survey estimated the populations receiving  
1146 collection by municipal utilities that owned or operated a sanitary sewerage system. The  
1147 purpose of the survey is to aid in overseeing and communicating with the wastewater industry in  
1148 Utah. We noticed that these data may have not be quantitatively accurate but are likely

1149 qualitatively informative. First, large swaths of the state appeared not to receive wastewater  
1150 collection (**Fig S 16**), but this may partly reflect that only 71% of contacts responding to the  
1151 survey [60]. Second, in some county subdivisions, >100% of the population appeared  
1152 connected to sewers, likely due to errors in estimation, differences in how combined treatment  
1153 and collection facilities reported population estimates, or collection populations resided in  
1154 neighboring county subdivisions. While the 5 county subdivisions with the highest percentages  
1155 of American Indian and Alaska Natives (~20-100%) were reported to not be connected to  
1156 sewers at all in this dataset, we believe this is because Indian reservations are not included in  
1157 the dataset (as they are under federal, not state jurisdiction). In fact, all 5 of these counties  
1158 (Uintah and Ouray, Casa del Eco Mesa-White Mesa, West Juab, Oljato, Blanding) include  
1159 Indian reservations. Additionally, the county subdivision with the highest percentage of Black or  
1160 African American; Native Hawaiian and Other Pacific Islander; or Hispanic are in county  
1161 subdivisions has nobody receiving collection (**Fig S 17**); however, we cannot rule out that this is  
1162 due to underreporting from these county subdivisions.

1163  
1164 The 2021 Minnesota Wastewater Infrastructure Needs Survey reported whether  
1165 communities have a collection system. The data from this survey was geographically  
1166 sparse, as communities (the level of aggregation of the dataset) often directly mapped  
1167 to a census designated place (CDP), and CDPs do not cover the entire state (**Fig S**  
1168 **14a**). Consequently, the sewer connectivity of the rural areas between census  
1169 designated places could not be determined from this dataset. Additionally, not all  
1170 communities responded to the survey. Of the 523 communities that participated in the  
1171 survey, all but 11 had a collection system. The geographic distribution of sewer  
1172 connectivity qualitatively matched that of the 2017 Subsurface Sewage Treatment  
1173 Systems (SSTS) Annual Report of the number of reported subsurface sewage

1174 treatment systems (septic tanks) by county in Minnesota (**Fig S 14b**); however, 7 out of  
1175 the 218 contacted local government units did not respond, 14 reported having zero  
1176 SSTS within their jurisdiction despite permitting SSTS and not all SSTS programs were  
1177 able to be identified to be contacted.

1178  
1179 Given the data incompleteness and biases for the Utah and Minnesota datasets, we  
1180 excluded them from the analyses in the main text.

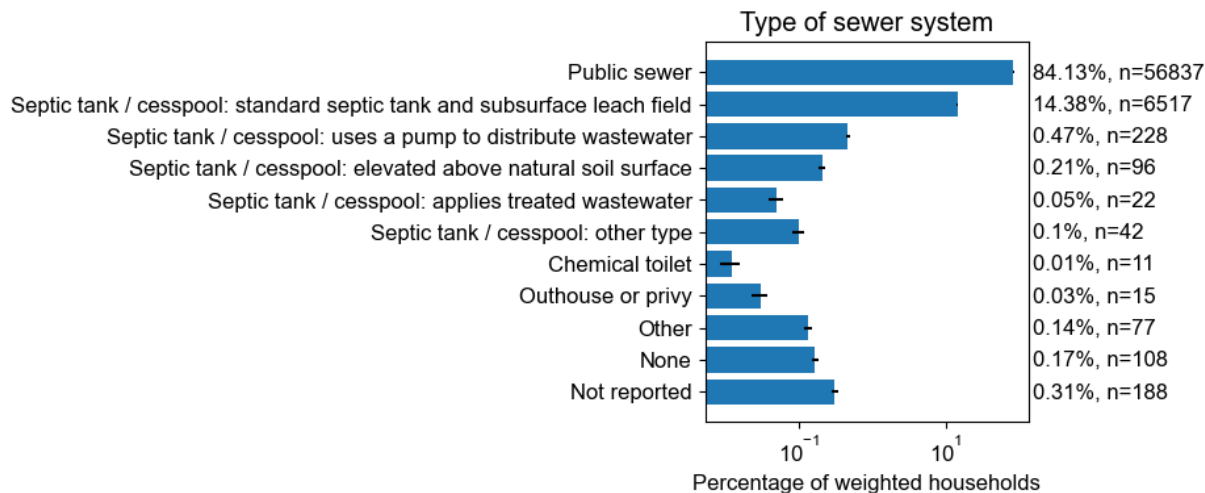
1181

## 1182 **Sewer connectivity in the U.S. Island Areas**

1183 From the 2020 US Census Island Areas Decennial Survey of household characteristics,  
1184 the overall levels of household sewer connection in the island areas were lower than in  
1185 the states (**Fig S 18**). In Guam, the Northern Mariana Islands, the Virgin Islands, and  
1186 American Samoa, the overall connection to sewer systems across each island was  
1187 65%, 51%, 66%, and 52% of households, respectively (compared to an overall 83%  
1188 connection across mainland US). To ask whether there was strong spatial variability  
1189 across each island, we looked at the median household connection by Census  
1190 Designated Place (CDP). A median of 64%, 32%, 70%, and 13% households by CDP in  
1191 Guam, the Northern Mariana Islands, the Virgin Islands, and American Samoa,  
1192 respectively, were connected to sewers. This suggests that sewer connectivity is  
1193 particularly spatially concentrated in the Northern Mariana Islands and American Samoa  
1194 (difference in median household connection by CDP and overall island connection),  
1195 although all island areas showed some spatial variability in connectivity when visually  
1196 assessing maps (**Fig S 19**). The only significant correlation of sewer connectivity with

1197 demographic or economic characteristics was in American Samoa, where CDPs with a  
1198 higher percentage of Black or African Americans were more connected to sewers  
1199 (Pearson  $r = 0.38$ ,  $q$ -value = 0.01).

## 1200 Supplementary figures



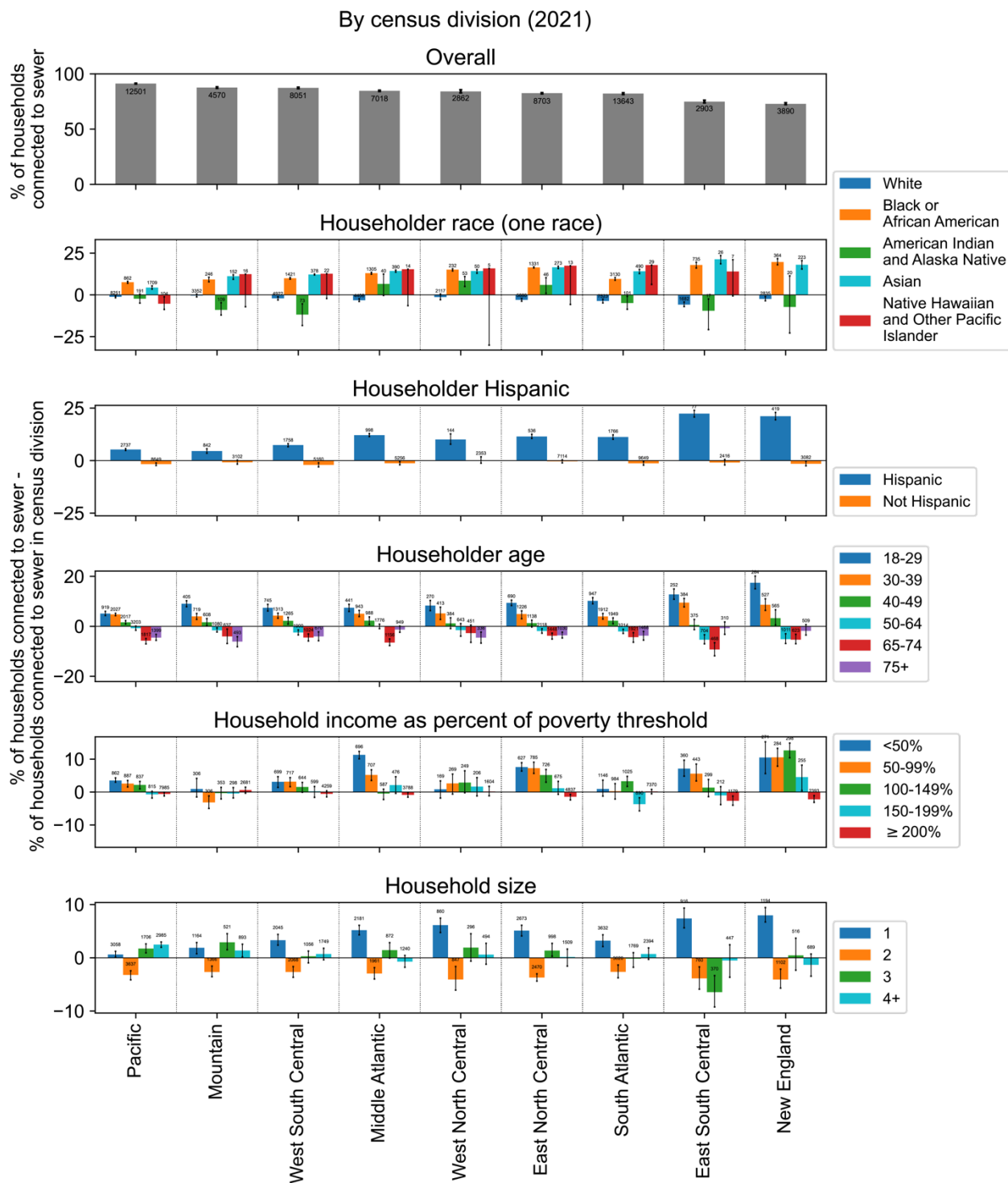
1201

1202 **Fig S 1: Percentage of households connected to various wastewater processing systems.**

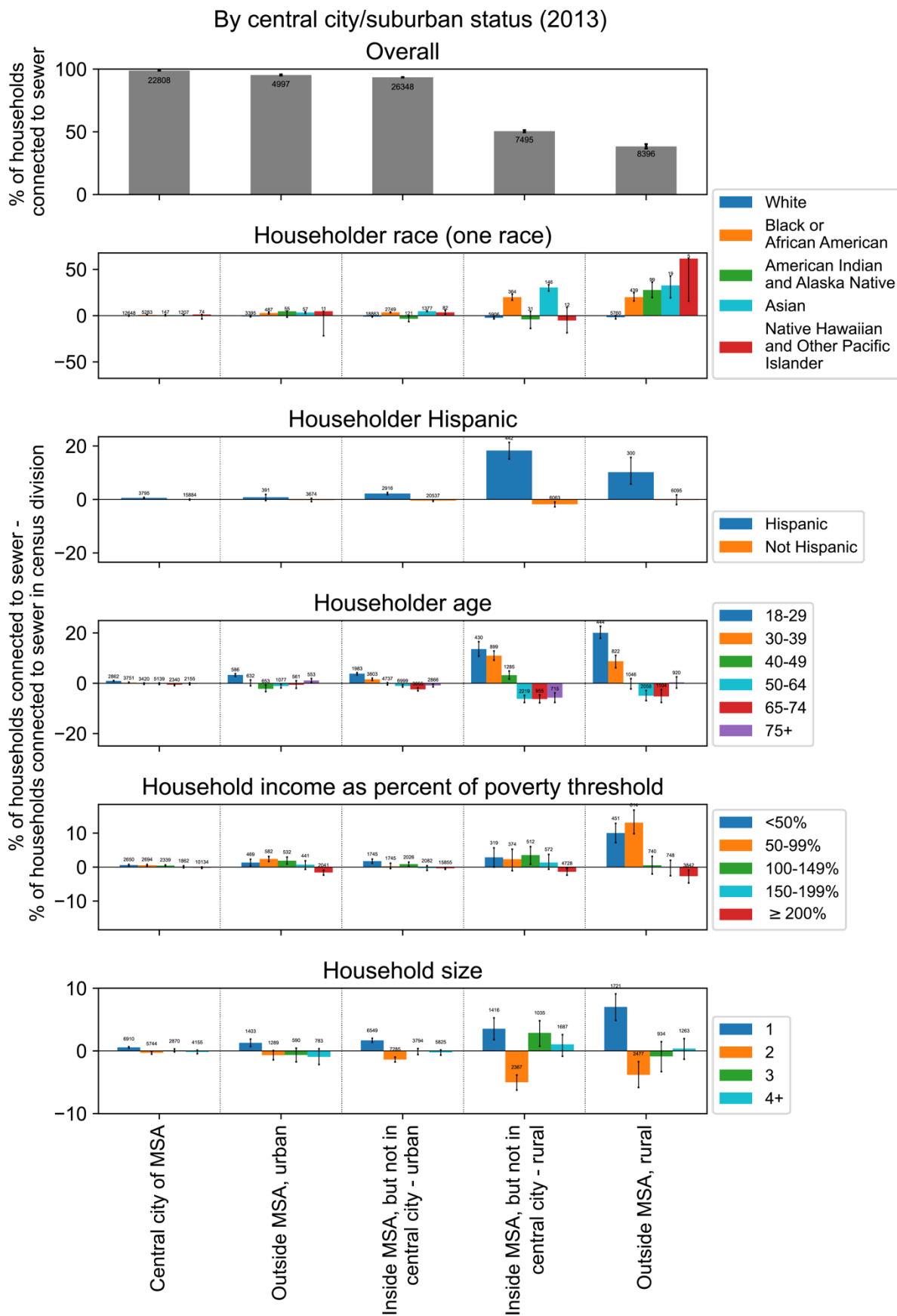
1203 The number of households sampled (n) for each category are also shown. Data are from 2021

1204 American Housing Survey.



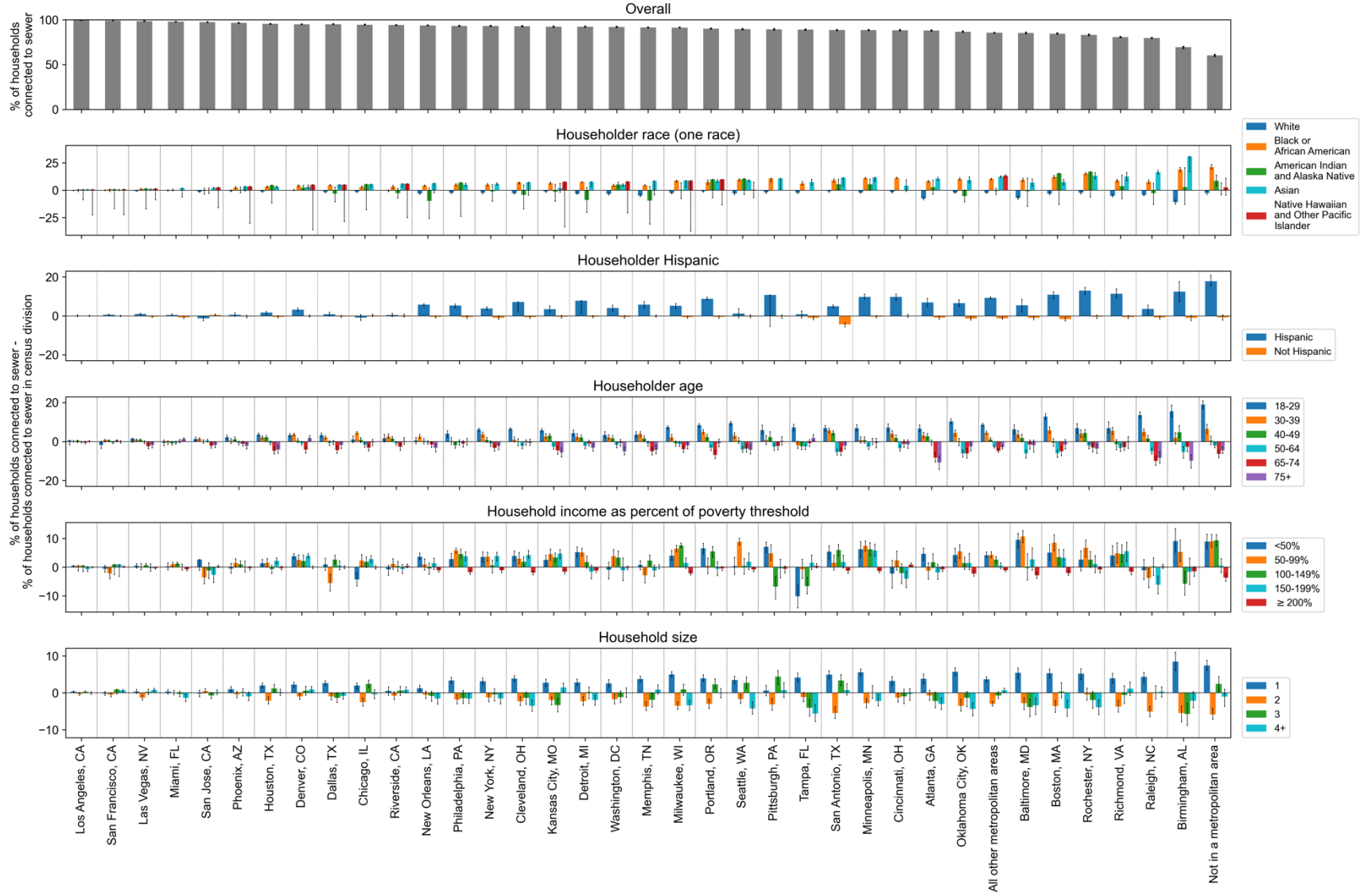


1205  
 1206 **Fig S 2: The percentage of households connected to sewers in each census division by**  
 1207 **demographic and economic characteristics (same as Fig 2) including the number of**  
 1208 **sampled households in each category.**



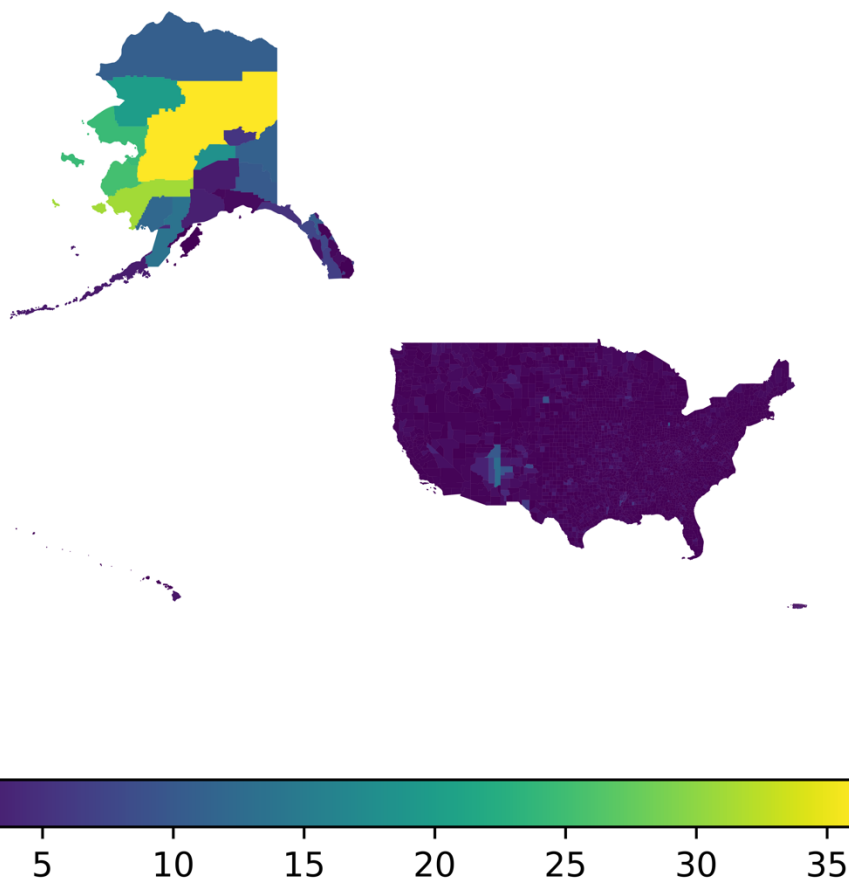
1210 **Fig S 3: The percentage of households connected to sewers by urban and rural**  
1211 **categorizations and by demographic and economic characteristics (Fig 3) including the**  
1212 **number of sampled households in each category.**

By core-based statistical areas (2019, 2021)



1214 **Fig S 4: Top 35 metropolitan areas' connectivity to public sewers by demographics.** Data are from the 2019 and 2021 U.S.  
1215 Census American Housing Survey [12,23]. The number of sampled households for each category can be found in the Github  
1216 repository in outputs/ahs\_sewer\_connectivity\_by\_cbsa.csv in the "num\_observations" column.

## Percent of occupied housing units lacking complete plumbing facilities by county

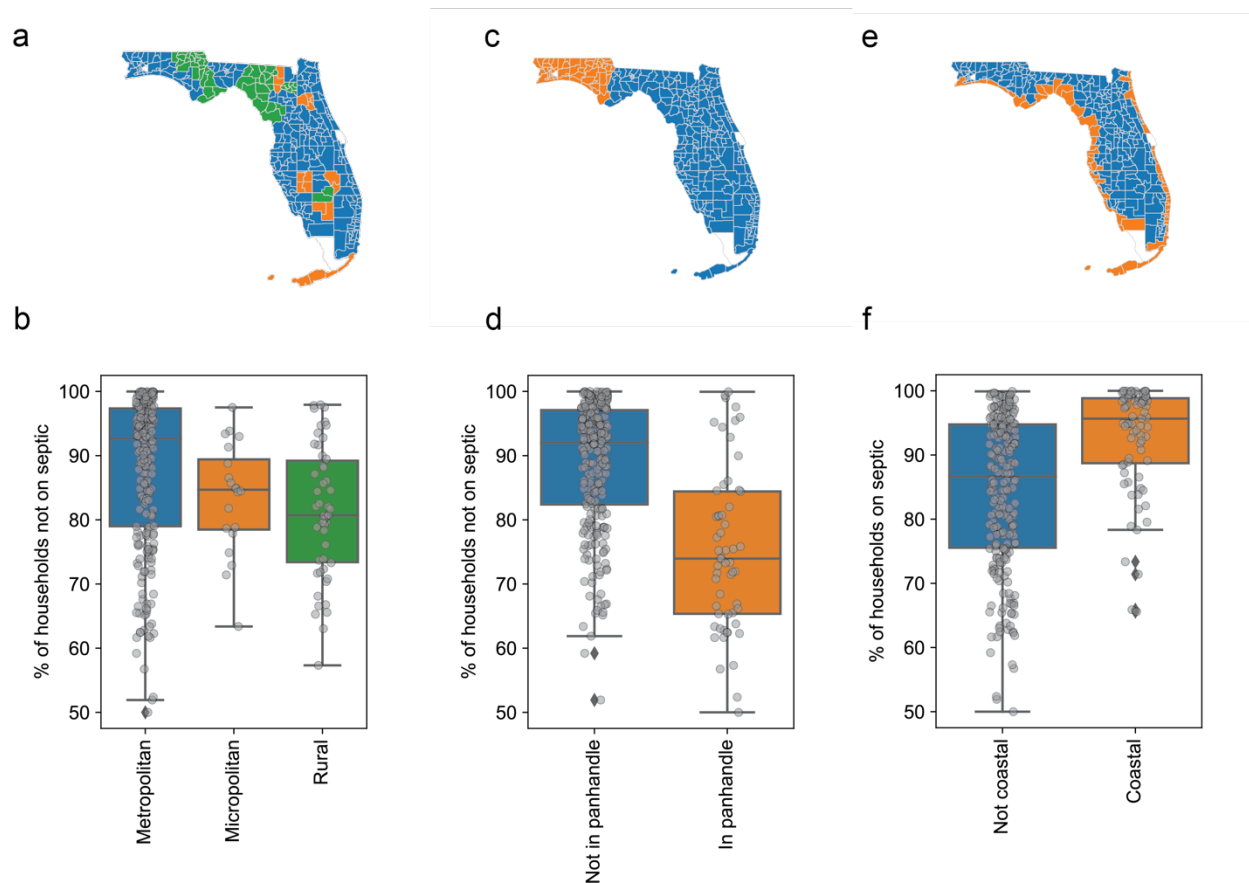


1217

1218

1219

**Fig S 5: Percent of occupied housing units lacking complete plumbing facilities by county.** Data are from the 2021 U.S. Census American Community Survey.



1220

1221 **Fig S 6: Comparison of percentage of households not on septic across different**

1222 **geographies in Florida.** Data are from the Florida Department of Health septic tank inspection

1223 permits reported in 2012. (a) Map of county subdivision colored by whether it is a Metropolitan

1224 Statistical Area (blue), Micropolitan Statistical Area (orange), or Rural Area (neither metropolitan

1225 not micropolitan, green). (b) Percentage of households not on septic in county subdivisions

1226 stratified by whether the county subdivision is a Metropolitan Statistical Area, Micropolitan

1227 Statistical Area, or Rural Area. County subdivisions that are Metropolitan Statistical Areas have

1228 a higher percentage of households not on sewer than those that are Micropolitan Statistical

1229 Areas, but the effect is not significant ( $p = 0.052$ ). County subdivisions that are Micropolitan

1230 Statistical Areas have a higher percentage of households not on sewer than those that are

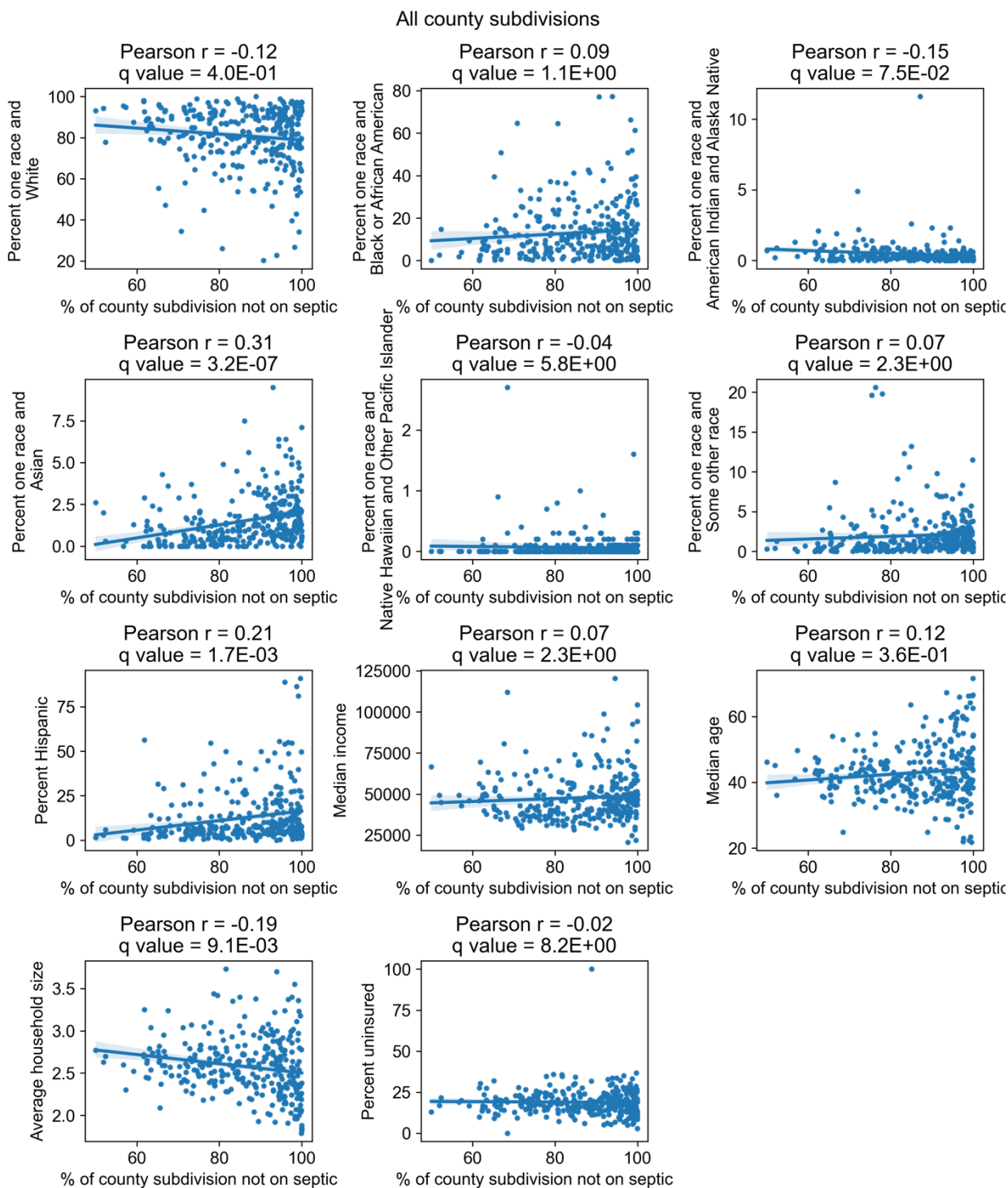
1231 Rural Areas, but the effect is not significant ( $p = 0.37$ ). (c) Map of county subdivisions colored by

1232 whether they are always included in references to the Florida panhandle. Orange: county

1233 subdivisions in the panhandle. Blue: county subdivisions not in the panhandle. (d) Percentage of  
1234 households not on septic in county subdivisions stratified by whether the county subdivision is in  
1235 the panhandle. Each point represents a county subdivision. County subdivisions in the  
1236 panhandle have a significantly lower percentage of households not on septic ( $p < 10^{-9}$ ). (e) Map  
1237 of county subdivisions colored by whether they have a coastline. Orange: county subdivision  
1238 with coastline. Blue: county subdivisions without coastline. (f) Percentage of households not on  
1239 septic in county subdivisions stratified by coastal and not coastal county subdivisions. Each  
1240 point represents a county subdivision. County subdivisions not on the coast have a significantly  
1241 higher fraction of households on septic ( $p < 10^{-9}$ ).



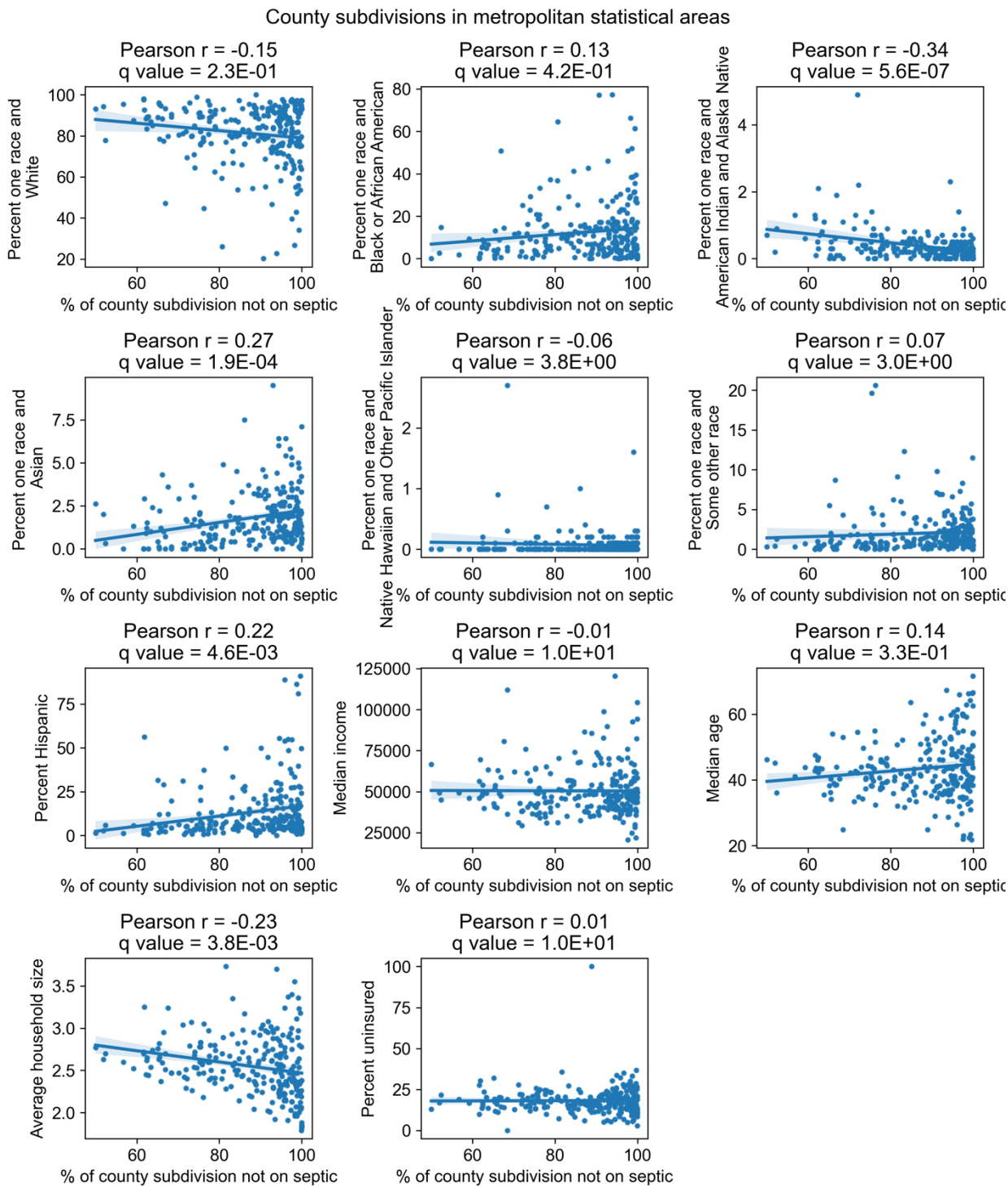
1242



1243

1244

1245 **Fig S 7: Correlation of percentage of households in county subdivisions of Florida *not***  
1246 **connected to septic tanks with demographic and economic factors.** Data are from the  
1247 Florida Department of Health septic tank inspection permits reported in 2012. Each point  
1248 represents a county subdivision. Only county subdivisions with at least 5 households and 20  
1249 population size were included.  
1250



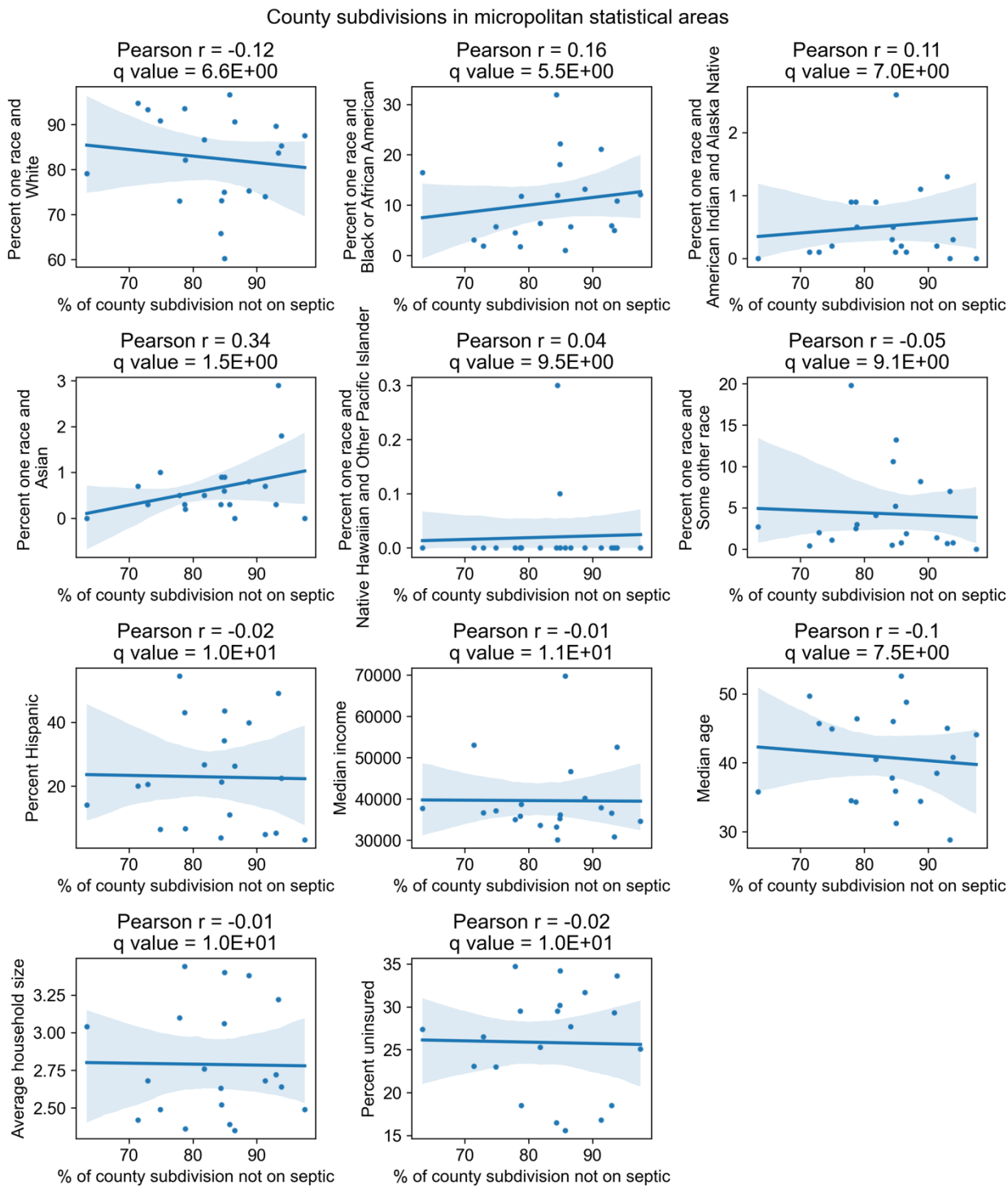
1251

1252 **Fig S 8: Correlation of percentage of households in county subdivisions of Florida in**

1253 **metropolitan statistical areas not connected to septic tanks with demographic and**

1254 **economic factors. Data are from the Florida Department of Health septic tank inspection**

- 1255 permits reported in 2012. Each point represents a county subdivision. Only metropolitan county  
1256 subdivisions with at least 5 households and 20 population size were included.



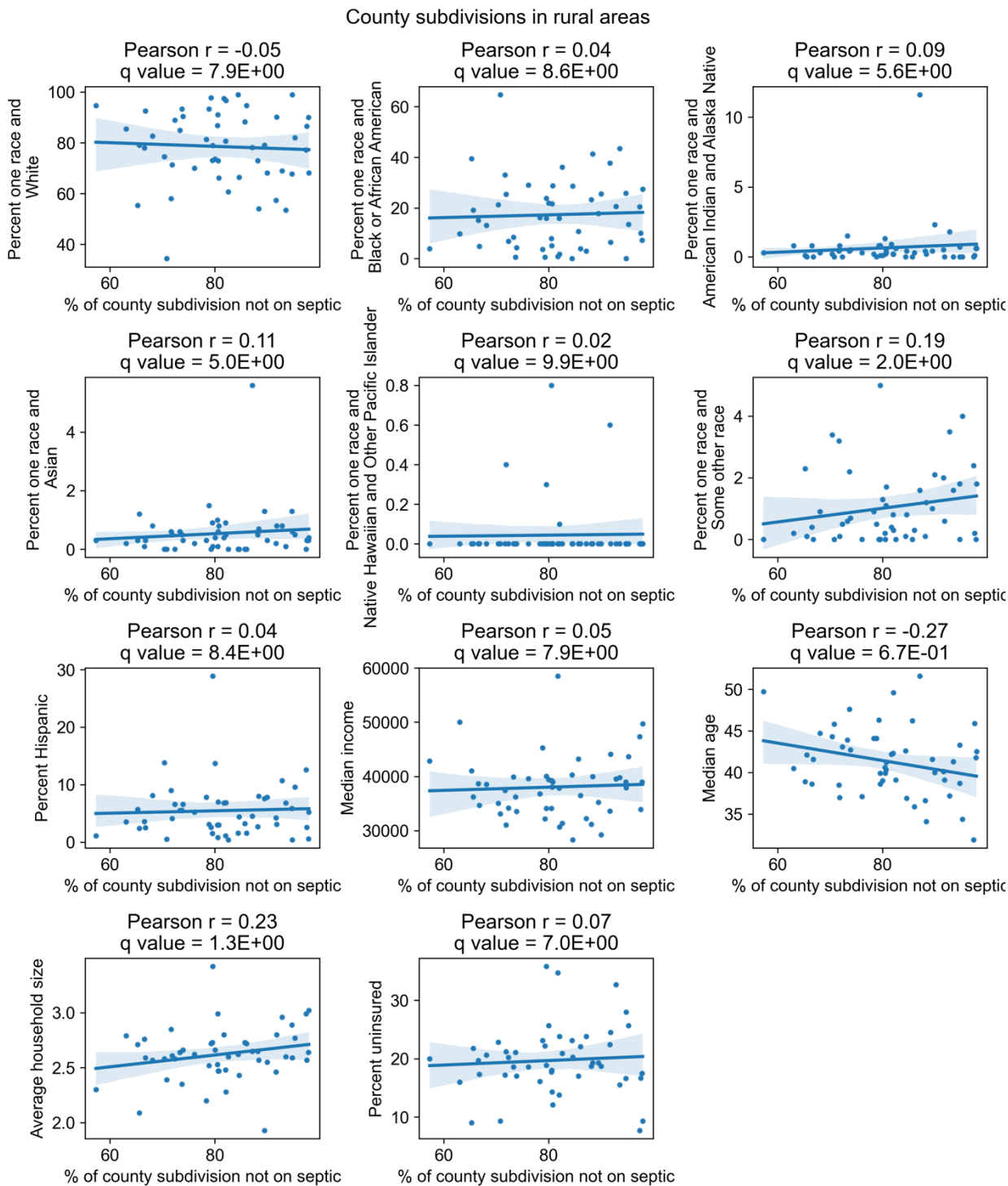
1257

1258 **Fig S 9: Correlation of percentage of households in county subdivisions of Florida in**

1259 **micropolitan statistical areas not connected to septic tanks with demographic and**

1260 **economic factors. Each point represents a county subdivision. Data are from the Florida**

- 1261 Department of Health septic tank inspection permits reported in 2012. Only micropolitan county  
1262 subdivisions with at least 5 households and 20 population size were included.



1263

1264

**Fig S 10: Correlation of percentage of households in county subdivisions of Florida in**

1265

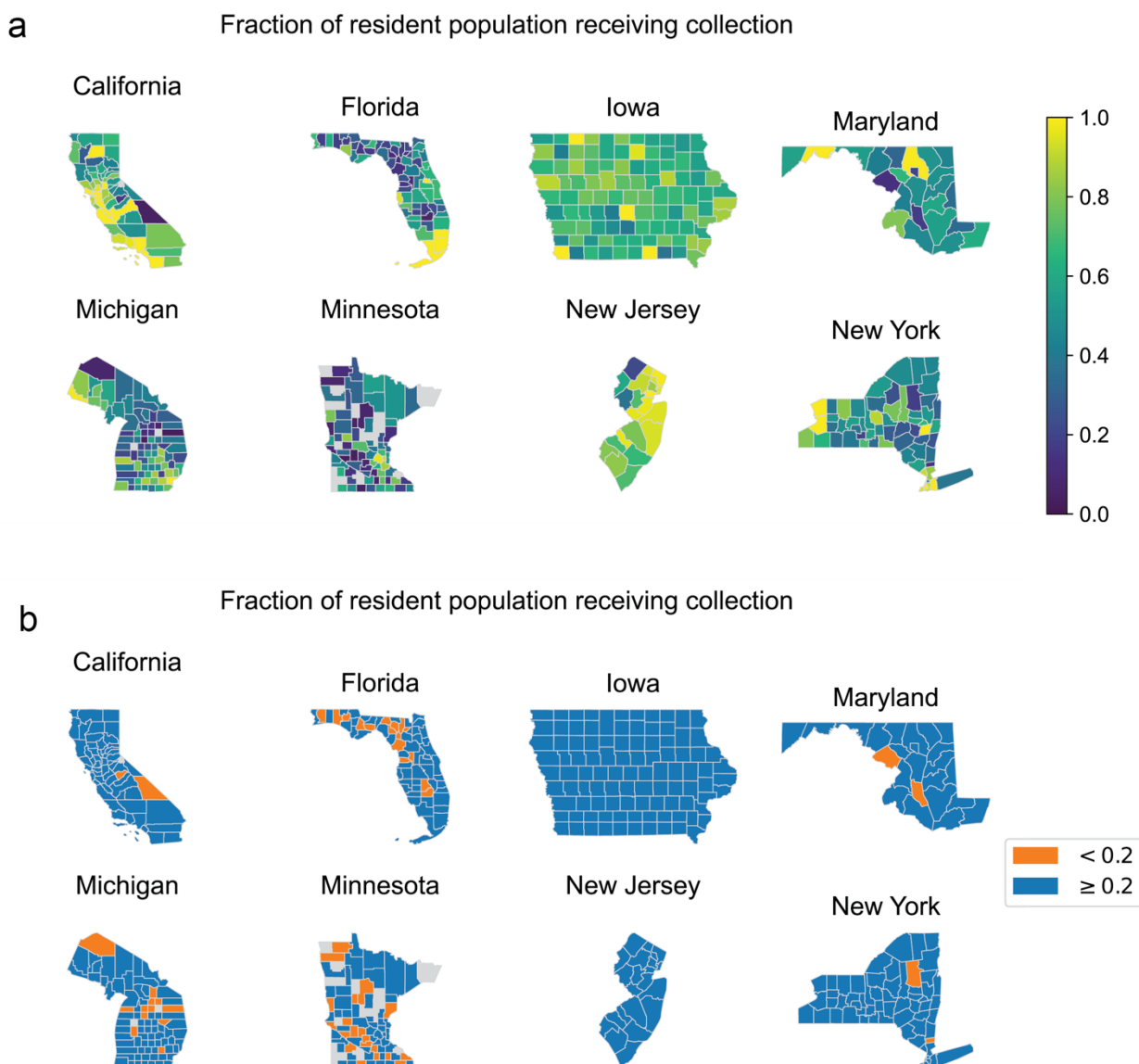
**rural areas not connected to septic tanks with demographic and economic factors. Data**

1266

are from the Florida Department of Health septic tank inspection permits reported in 2012. Each

1267 point represents a county subdivision. Only rural county subdivisions with at least 5 households  
1268 and 20 population size were included.

1269  
1270



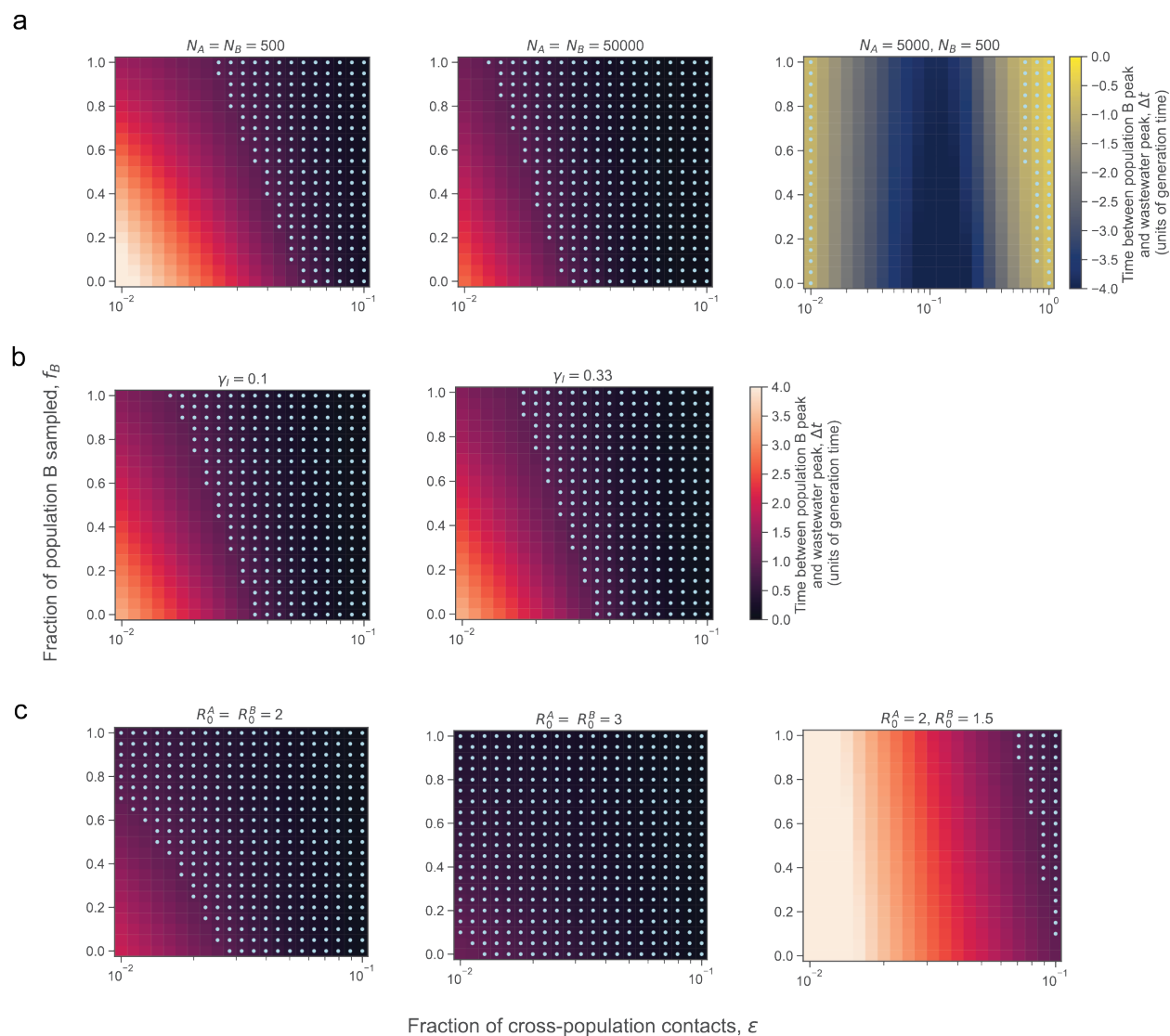
1271  
1272 **Fig S 11: Fraction of present resident population receiving collection by county in 8**  
1273 **states.** Data are from the 2012 EPA Clean Watersheds Needs Survey. (a) Fraction shown on a  
1274 continuous scale. (b) Fraction above and below 0.2. Only states that had more comprehensive  
1275 responses in the survey are shown. Counties that did not report any data in the survey are



1276 shown in gray. Counties that report more than 100% of the present resident population  
 1277 receiving collection are shown in (a) as 1 and in (b) as  $\geq 0.2$ .

1278

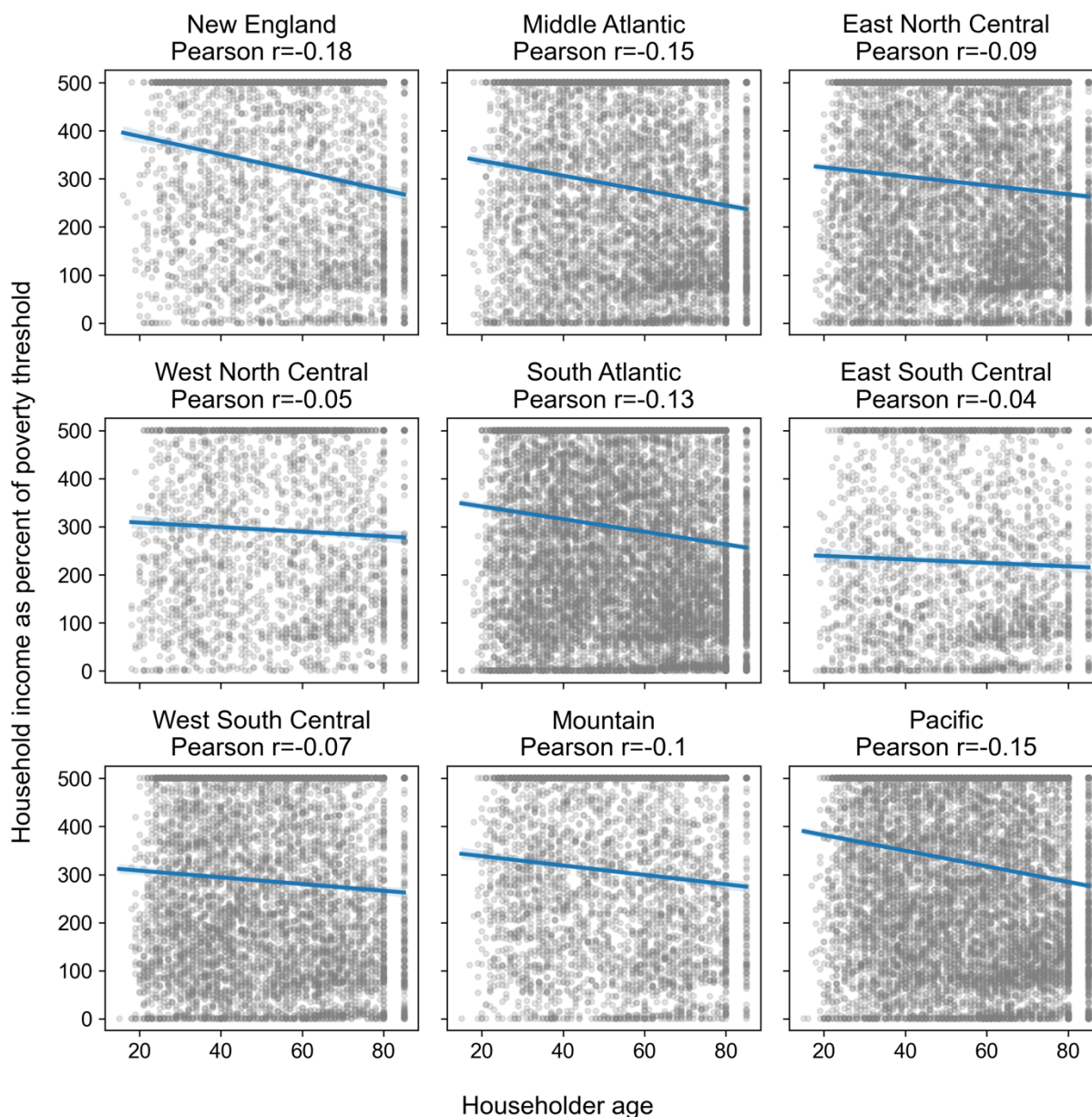
1279



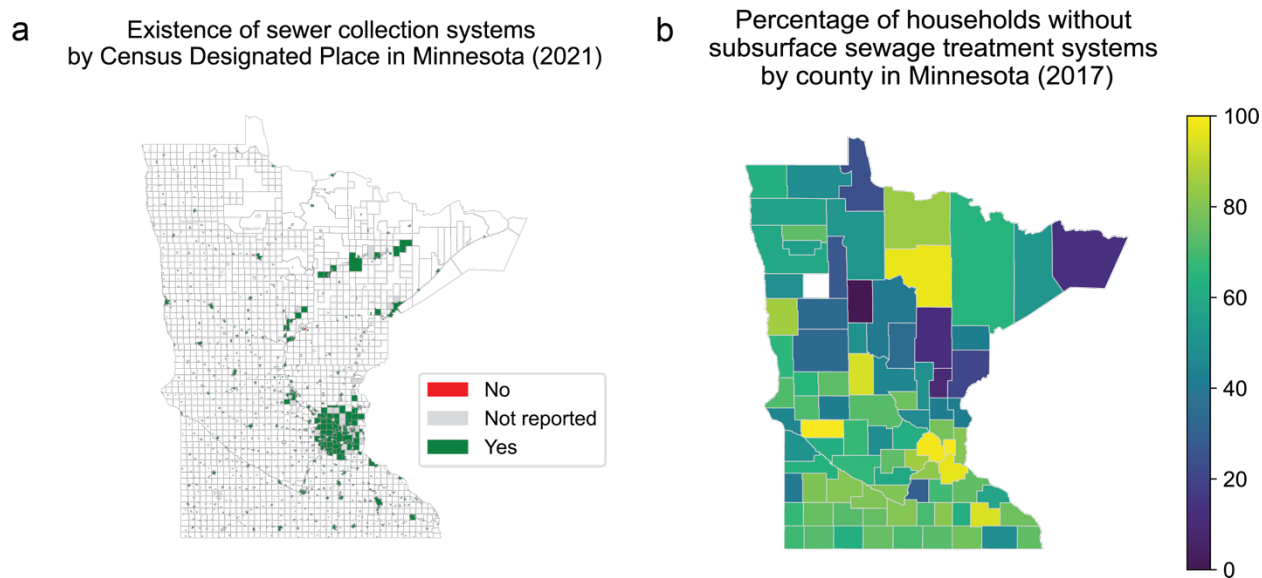
1280

1281 **Fig S 12: Sensitivity analysis of Fig 5c from main text.** The effect on the time between  
 1282 population B peak and wastewater peak of varying the (a) population size of the two populations  
 1283 ( $N_A$ ,  $N_B$ ), (b) recovery rate ( $\gamma_i$  in units of inverse days), and (c) basic reproduction number of the

1284 two populations ( $R_0^A$ ,  $R_0^B$ ) across a range of values. Unless where indicated, the parameters  
1285 were  $N_A=N_B=5000$ ,  $\gamma_i=0.18$  days<sup>-1</sup>, and  $R_0^A$ ,  $R_0^B=1.5$ .

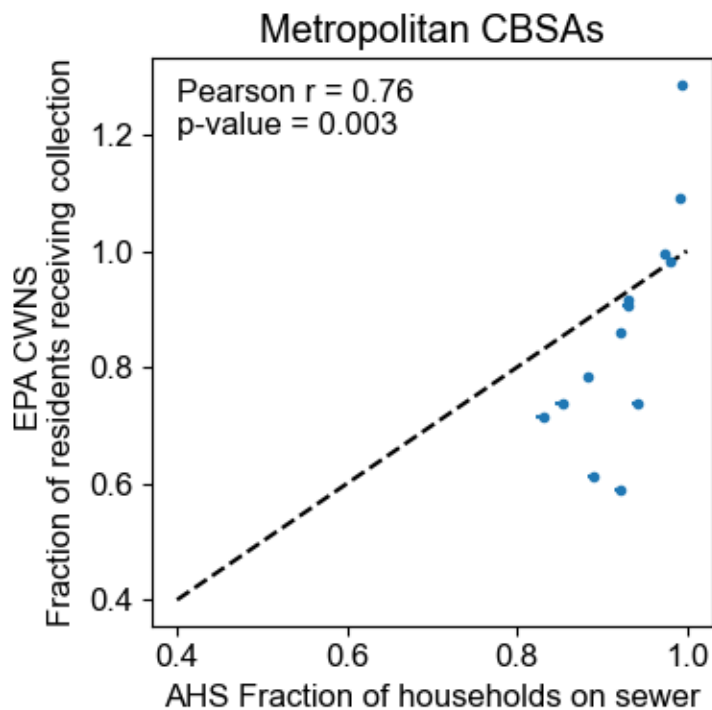


1286  
1287 **Fig S 13: Correlation between householder age and household income in the American**  
1288 **Housing Survey dataset by census division.**



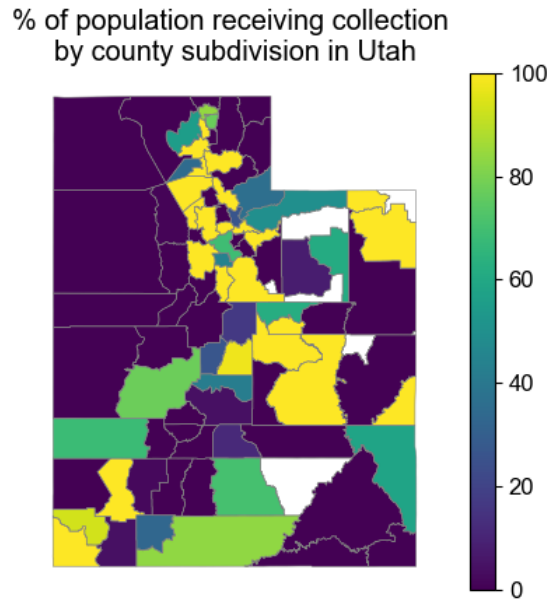
1289

1290 **Fig S 14: Sewer connectivity in Minnesota.** (a) Map showing whether a community in  
1291 Minnesota (mapped to Census Designated Place) reported having a collection system in the  
1292 2021 Minnesota Wastewater Infrastructure Needs Survey. Almost all communities mapped to  
1293 Census Designated Places. (b) Percentage of households without subsurface sewage treatment  
1294 systems (i.e. without septic tanks, suggesting sewerage) by county reported in the 2017  
1295 Subsurface Sewage Treatment Systems in Minnesota Annual Report [62]. Counties reporting  
1296 more than 100% of households having subsurface sewage treatment systems are set at 100%  
1297 (shown as 0% without subsurface sewage treatment systems in the map).



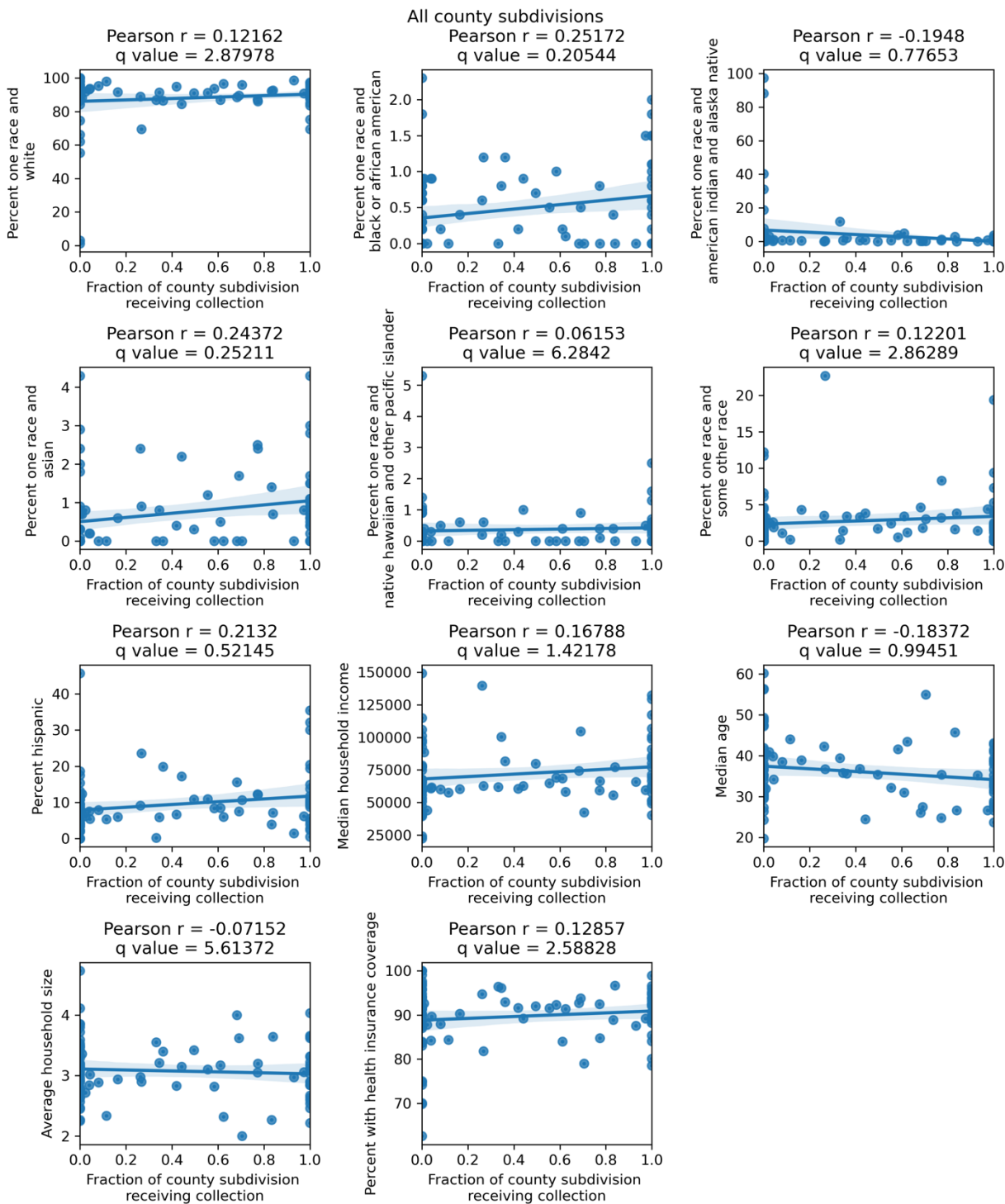
1298

1299 **Fig S 15: Comparison of the 2012 EPA Clean Watersheds Needs Survey (CWNS) and the**  
1300 **2019 and 2021 U.S. Census American Housing Survey (AHS) datasets in core based**  
1301 **statistical areas (CBSA).** Data from the EPA CWNS are of the fraction of residents in the  
1302 CBSA receiving sewage collection. Data from the AHS are of the fraction of households in the  
1303 CBSA on sewer. Only CBSAs that were oversampled in the AHS data and were in states that  
1304 had more comprehensive responses in the EPA CWNS were included (see Methods). The  
1305 dashed black line shows  $y=x$  as a reference.



1306

1307 **Fig S 16: The percentage of the population that received wastewater collection by a**  
1308 **municipal utility in Utah by county subdivision of Utah.** Data are from the 2021 Utah  
1309 Municipal Wastewater Planning Survey. All values less than 0 are displayed as 0. County  
1310 subdivisions with fewer than 5 households or 20 population size are displayed in white. Note  
1311 that Indian reservations were not surveyed in this dataset.

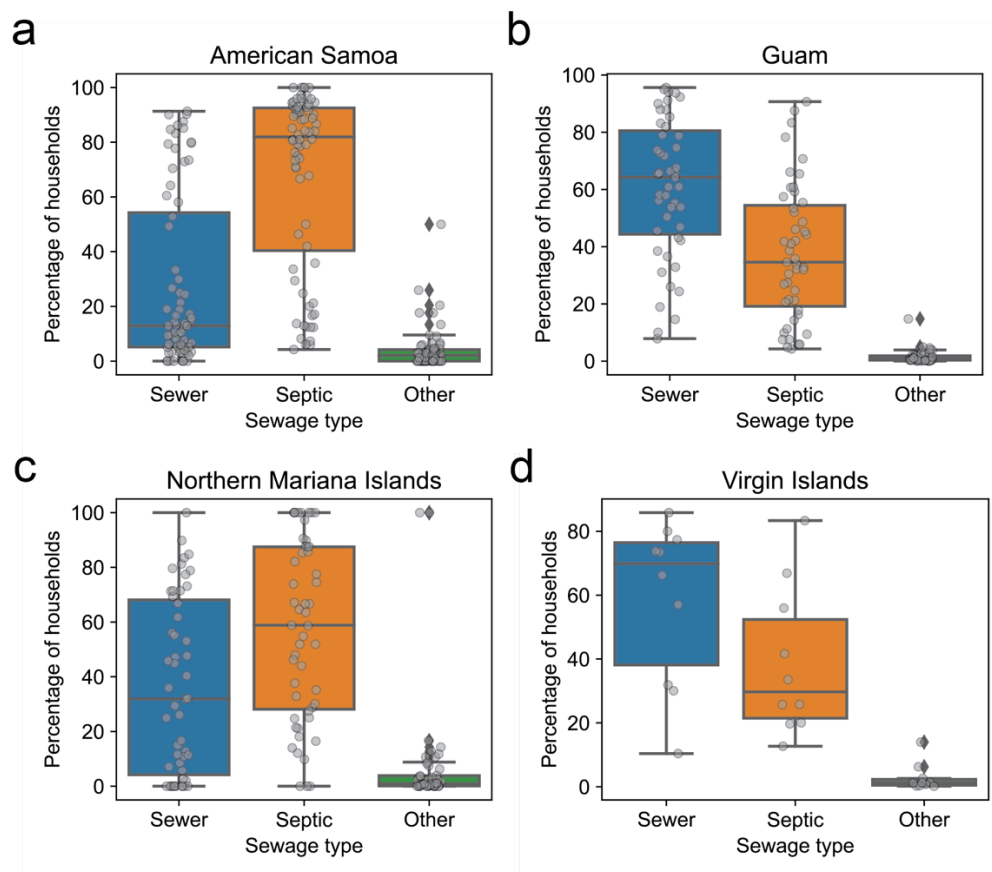


1312

1313 **Fig S 17: Correlation of fraction households in county subdivisions of Utah with**  
 1314 **wastewater collection with demographic and economic factors.** Data are from the 2021  
 1315 Utah Municipal Wastewater Planning Survey. Note that Indian reservations were not surveyed

1316 in this dataset. Values of the fraction of the county subdivision receiving collection greater than  
1317 1 were set to 1.

1318

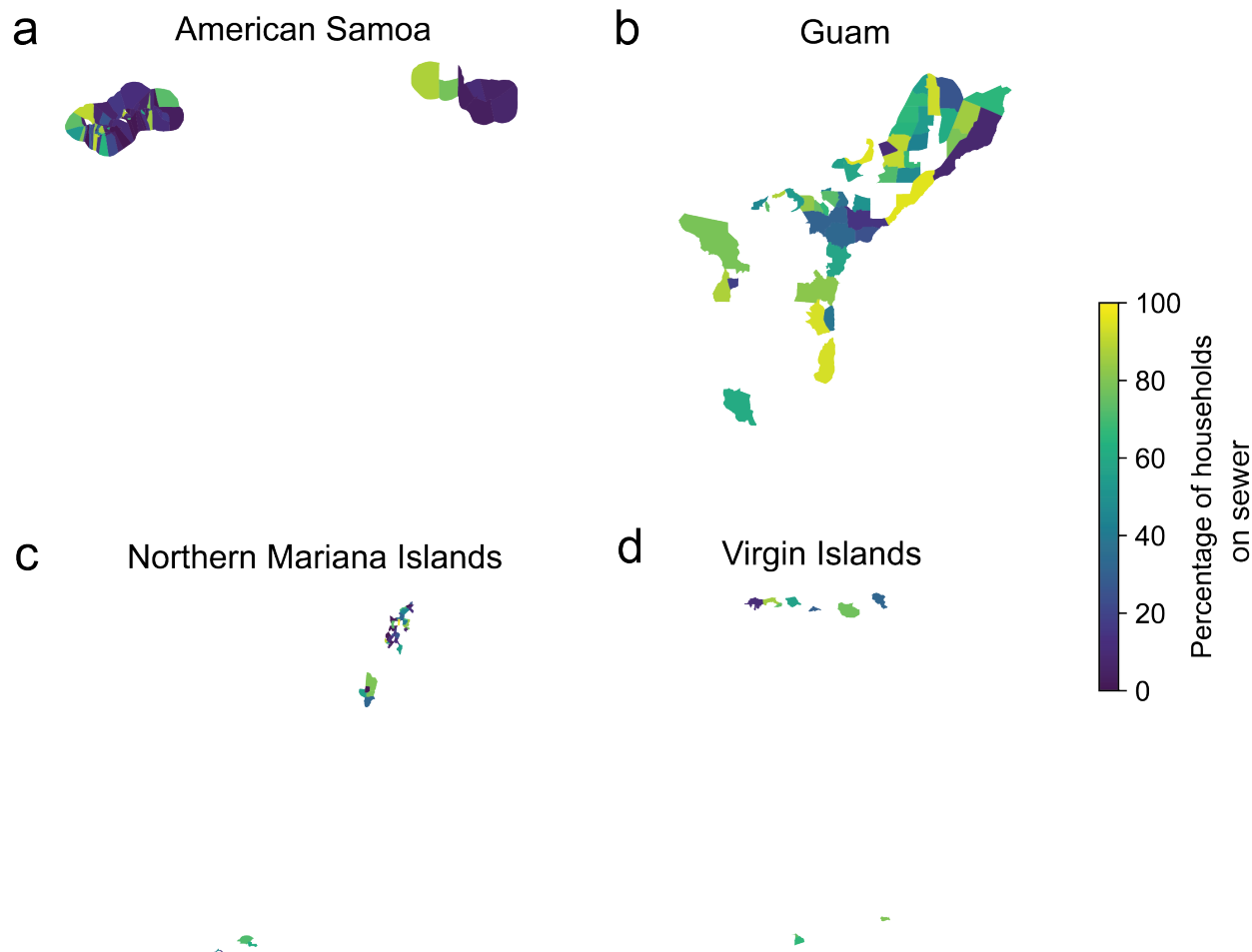


1319

1320 **Fig S 18: Fraction of households by Census Designated Place that are connected to**  
1321 **sewers, septic tanks, or other forms of sewage disposal in the U.S. Island Areas. (a)**

1322 American Samoa, (b) Guam, (c) the Northern Mariana Islands, and (d) the Virgin Islands. Data

1323 are from the 2020 U.S. Census Island Areas Decennial Survey.



1324

1325 **Fig S 19: Map of fraction of households connected to sewers by Census Designated**

1326 **Place.** (a) American Samoa, (b) Guam, (c) the Northern Mariana Islands, and (d) the Virgin

1327 Islands. Data are from the 2020 U.S. Census Island Areas Decennial Survey.

1328

1329



1330 **Supplementary tables**

1331 **Table S 1: Counties in the US with  $\geq 5\%$  of occupied housing units lacking complete**

1332 **plumbing facilities.** Data are from the 2021 U.S. Census American Community Survey.

	<b>County, State</b>	<b>Percent of occupied housing units lacking complete plumbing facilities</b>
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

1333

1334 **Table S 2: Correlation of the percentage of households in a county connected to**  
 1335 **public sewers with different demographic or economic factors by state; same as**  
 1336 **Table 3, but including all state-variable pairs.**

	Correlation of variable		
Demographic or economic variable	State	with % of households by county connected to public	Significance (q-value)

sewers (Pearson correlation  
coefficient)

Average household size	MN	0.29	0.49
	IA	-0.24	1
	NJ	0.34	1
	MD	-0.32	1
	NY	0.19	1
	FL	-0.12	1
	CA	0.25	1
	MI	0.01	1
Median age	MN	-0.43	<0.01
	MI	-0.41	0.01
	CA	-0.41	0.13
	NY	-0.37	0.26
	MD	0.17	1
	IA	-0.08	1
	NJ	-0.42	1
	FL	-0.05	1
Median income	MN	0.44	<0.01
	FL	0.48	<0.01
	CA	0.49	<0.01
	NJ	-0.27	1
	IA	-0.11	1

	MD	-0.17	1
	MI	0.21	1
	NY	0.23	1
	FL	0.36	0.26
	NY	0.31	1
	NJ	0.51	1
	IA	0.15	1
Percent Hispanic	MD	-0.29	1
	MI	0.24	1
	CA	0.25	1
	MN	0.19	1
	CA	-0.50	<0.01
	FL	-0.29	1
	MN	-0.14	1
Percent one race and American Indian and Alaska Native	NJ	0.09	1
	NY	0.03	1
	MI	-0.02	1
	MD	-0.04	1
	IA	-0.01	1
	MN	0.58	<0.001
Percent one race and Asian	MI	0.51	<0.001
	NY	0.55	<0.001
	CA	0.56	<0.001

	FL	0.53	<0.001
	IA	0.22	1
	NJ	0.25	1
	MD	-0.02	1
	MN	0.62	<0.001
	MI	0.59	<0.001
	NY	0.55	<0.001
Percent one race and Black or African American	IA	0.38	<0.01
	CA	0.46	0.02
	NJ	0.39	1
	MD	-0.05	1
	FL	0.01	1
	IA	0.26	0.85
	MN	-0.18	1
	MI	-0.15	1
Percent one race and Native Hawaiian and other Pacific Islander	MD	0.28	1
	NY	0.10	1
	CA	0.07	1
	NJ	-0.07	1
	FL	-0.01	1
Percent one race and some other race	NJ	0.57	0.63
	CA	0.35	0.68
	MN	0.26	1

	NY	0.23	1
	IA	0.16	1
	MI	0.17	1
	MD	-0.28	1
	FL	0.03	1
	MI	-0.57	<0.001
	CA	-0.54	<0.001
	NY	-0.51	<0.01
Percent one race and	IA	-0.35	0.03
White	MN	-0.28	1
	NJ	-0.54	1
	FL	-0.08	1
	MD	0.08	1
	MI	-0.38	0.03
	NJ	0.51	1
	CA	-0.24	1
Percent uninsured (health	MN	-0.15	1
insurance)	NY	-0.10	1
	MD	-0.15	1
	FL	0.06	1
	IA	-0.03	1

1337 q-values above 1 are set to 1.

1338 **Table S 3: Correlation of the fraction of households by county connected to public**  
 1339 **sewers and different demographic or economic factors, stratified by state and**  
 1340 **Metropolitan, Micropolitan, and Rural Statistical Areas.** Only correlations that are significant  
 1341 (q-value<=0.05) are shown.

Demographic or economic variable	State	Metropolitan, Micropolitan, or Rural	Pearson	
		Statistical Area	r	q-value
		Micropolitan Statistical Area		
Median income	FL	Area	0.98	9.8E-03
		Metropolitan Statistical Area		
Percent Hispanic	FL	Area	0.53	1.8E-02
		Metropolitan Statistical Area		
	MN	Area	0.72	2.3E-03
Percent one race and Asian	FL	Area	0.52	3.0E-02
		Metropolitan Statistical Area		
Percent one race and Black or African American	MN	Area	0.73	1.1E-03
		Metropolitan Statistical Area		
American	MI	Area	0.67	1.4E-02
		Metropolitan Statistical Area		
	MN	Area	-0.74	7.6E-04
Percent one race and White	MI	Area	-0.71	4.6E-03

### Metropolitan Statistical

IA	Area	-0.69	5.0E-02
----	------	-------	---------

1342

1343



1344

1345 **Table S 4: Additional datasets discussed in the Supplementary Information but excluded**

1346 **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 [60]. 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.	[60]
Minnesota Wastewater Infrastructure Needs Survey (WINS)	Minnesota	2021	Communities (places and townships) responded 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.	[61]
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 [62], "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.	[62]

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

1347

1348 **Table S 5: Manual changes community names reported in the Minnesota**

1349 **Wastewater Infrastructure Needs Survey to be able to match with a U.S. Census**

1350 **Designated Place.**

<b>Original name in dataset</b>	<b>Edited name for merging with Census Designated Place name</b>	<b>Note</b>
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.