

1 **Assessing health equity in wastewater monitoring programs: Differences in the demographics and**
2 **social vulnerability of sewerred and unsewered populations across North Carolina**

3 Xindi C. Hu^{1,*†}, Stacie K. Reckling^{2,3,*}, Aparna Keshaviah¹

4 ¹ Mathematica, Inc., Princeton, New Jersey, 08543, USA

5 ² North Carolina Department of Health and Human Services, Division of Public Health, Raleigh,
6 North Carolina, 27609, USA

7 ³ Center for Geospatial Analytics, North Carolina State University, Raleigh, North Carolina 27695, USA

8 *Xindi C. Hu, Stacie Reckling: These authors contributed equally to this work

9 †*Correspondence to:* Xindi C. Hu, 600 Alexander Park Dr, Princeton, NJ 08540. Telephone: +1-
10 5102854675. Email: chu@mathematica-mpr.com

11

12 Conflicts of Interest:

13 The authors declare they have nothing to disclose.

14 **Abstract**

15 **Background:** Wastewater monitoring is a valuable tool to track community-level disease trends.

16 However, the extent to which vulnerable populations have been included in statewide wastewater
17 monitoring programs remains unstudied.

18 **Objectives:** We conducted a geospatial analysis to examine (1) the representativeness of wastewater data
19 collected through the North Carolina Wastewater Monitoring Network as of June 2022, and (2) the
20 potential of wastewater data to generalize to unsewered populations in the county.

21 **Methods:** After intersecting wastewater treatment plant (WWTP) service areas (sewersheds) with census
22 block and tract boundaries for 38 WWTPs across 18 counties, we compared the demographics and social
23 vulnerability of (1) people residing in sewersheds of monitored WWTPs with countywide and statewide
24 populations, and (2) people connected to any sewer system—regardless of inclusion in wastewater
25 monitoring—with unsewered populations. We flagged differences greater than +/- 5 percentage points or
26 percent (for categorical and continuous variables, respectively) and noted which were statistically
27 significant (i.e., greater than twice the margin of error).

28 **Results:** As a whole, populations in monitored sewersheds resembled the statewide population on most
29 demographics analyzed, with a few exceptions. When multiple WWTPs were monitored within a county,
30 their combined service populations resembled the countywide population, although populations in
31 individually monitored sewersheds sometimes differed from the countywide population. In nine counties
32 for which we had comprehensive sewershed maps, we found that sewer residents had higher social
33 vulnerability, a greater share of Hispanics and African Americans, lower income, and lower educational
34 attainment than unsewered residents.

35 **Discussion:** Our results suggest that wastewater monitoring in North Carolina well represents the larger
36 community. Ongoing analyses will be needed as sites are added or removed. The approach we present

- 37 here can be used to ensure that wastewater surveillance programs nationwide are implemented in a
- 38 manner that informs equitable public health decision-making.

39 **Introduction**

40 Early in the COVID-19 pandemic, clinical testing was restricted due to mass test kit shortages across the
41 United States. Access to testing—a critical public health resource—aligned with structural disparities,
42 with inequities among minority, uninsured, and rural groups (Rader et al. 2020). In poorer areas, there
43 were fewer testing sites per person, and those sites were located farther away (Kim 2020). Communities
44 that were majority Black and Hispanic were also more likely to face longer wait times and understaffed
45 testing centers.

46 Recognizing that a better way existed to monitor population-wide infection levels, hundreds of
47 communities launched wastewater testing for the SARS-CoV-2 virus that causes COVID-19. Wastewater
48 monitoring can cover a much broader swath of the population than clinical testing, and taps into existing
49 sanitation infrastructure, providing a practical and scalable solution to public health surveillance
50 (Keshaviah 2021). In the United States, 16,000 wastewater treatment plants capture sewage from roughly
51 75% of the population (US EPA Office of Water, 2015). Worldwide, researchers estimate that roughly 1
52 in 4 people is connected to a wastewater treatment plant (Hart and Halden 2020). Critically, wastewater
53 monitoring captures health biomarkers of sewered populations regardless of whether they visit a testing
54 site or doctor, and regardless of whether they have symptoms, since people with asymptomatic infections
55 also shed viral particles into their stools (Noh et al. 2020; Zhou et al. 2020).

56 Despite the potential of wastewater monitoring to improve health equity, resource constraints may inhibit
57 equitable access to this innovative approach to public health surveillance. Prior to COVID-19, wastewater
58 monitoring for diseases and controlled substances rarely occurred in low- or middle-income countries
59 (LMICs). Of the fourteen countries that had routinely employed environmental surveillance for poliovirus
60 as part of the Global Polio Eradication Initiative, ten (71%) were high-income countries (Hovi et al.
61 2012). Likewise, of the 37 member countries in the Sewage analysis CORE group—Europe network,
62 which coordinates international wastewater studies on drug use in and beyond Europe, only 5 (14%) are
63 LMICs (González-Mariño et al. 2020; World Bank n.d.). Even after the global expansion of wastewater

64 testing to help officials worldwide manage the coronavirus pandemic, research has shown that monitoring
65 has primarily occurred in high-income countries (Naughton et al. 2021).

66 While wastewater monitoring has the potential to overcome disparities in public health surveillance, little
67 research has been conducted to determine the comparability of sewered and unsewered populations with
68 respect to demographics and social vulnerability, and whether communities included in state and national
69 wastewater monitoring programs resemble to those not being monitored. The National Academies
70 Sciences, Engineering, and Medicine (2023) report, which stressed the importance of equity in national
71 wastewater monitoring efforts, implied that because many unsewered households are in rural areas, and
72 because rural areas tend to be more disadvantaged than urban areas, it follows that unsewered populations
73 are more disadvantaged than sewered populations. However, an analysis of data from the 2019 American
74 Household Survey found the opposite to be true—that septic households are more economically
75 *advantaged* than sewered households—with the pattern upheld even when analyses were stratified by
76 urbanicity (Olesen et al. 2022). Given that the Centers for Disease Control and Prevention’s National
77 Wastewater Surveillance System (CDC NWSS) will continue to fund state, local, and tribal wastewater
78 programs through at least 2025, officials need a way to ensure that wastewater sampling sites are selected
79 in a manner that promotes health equity.

80 To assess how representative wastewater data collected during the COVID-19 pandemic are and could be,
81 we conducted detailed geospatial analyses to answer two key questions: (1) Are sewered populations
82 monitored through wastewater surveillance representative of the counties they come from with respect to
83 demographics and social vulnerability? (2) How similar are the demographics and social vulnerability of
84 communities that are and are not connected to a sewer system (regardless of inclusion in a wastewater
85 monitoring program)? We focused our analysis on North Carolina, one of the first eight states funded by
86 CDC NWSS.

87 **Methods**

88 Study site selection

89 Across both sets of analyses, we analyzed the service populations of 38 WWTPs in 18 counties, including
90 25 actively monitored WWTPs as of June 2022, 1 previously monitored WWTP, and 12 WWTPs not
91 monitored by the NCWMN.

92 The North Carolina Wastewater Monitoring Network (NCWMN) collects wastewater samples from
93 municipal wastewater treatment plants (WWTPs) across the state twice weekly and analyzes these
94 samples to quantify SARS-CoV-2 viral concentrations. As of June 2022, the NCWMN routinely analyzed
95 samples from 25 WWTPs across 17 of 100 NC counties, covering roughly one-fourth of the state's
96 population. Sites were added to the state's wastewater monitoring program in stages. Because the initial
97 group of sites came from a COVID-19 wastewater surveillance pilot project coordinated by universities
98 (Noble et al. 2021), they were mainly located in urban centers near university labs that had the capacity
99 analyze wastewater samples. Over time, the NCWMN expanded to include sites in other areas of the
100 state, including the rural mountainous region in Western North Carolina, and underserved communities
101 with higher social vulnerability, low COVID-19 vaccination rates, or both (NCDHHS 2021). In addition,
102 five sites were added to the NCWMN when the Wake County health department initiated a local
103 wastewater monitoring program (Supplemental Materials (SM) Figure S1).

104 To assess the representativeness of wastewater data, we conducted two sets of analyses. First, we
105 analyzed the demographic and social vulnerability characteristics of people living in monitored
106 sewersheds versus the general population, comparing: (A) individual monitored sewershed populations
107 with their respective countywide population, (B) combined monitored sewershed populations, aggregated
108 to the county level (in the case that multiple WWTPs were monitored within a county), with the
109 respective countywide population and (C) combined monitored sewershed populations, aggregated to the
110 state level, with the statewide population. Second, we compared the demographics and social
111 vulnerability of sewerred and unsewerred populations in nine counties for which we could wholly identify
112 the county's sewerred population using geospatial shapefiles delineating the service areas of all municipal
113 WWTPs in the county with a treatment capacity of more than 0.5 million gallons per day. The nine

114 counties covered rural and non-rural counties from across the state and included eight counties actively
115 participating in NCWMN as of June 2022 plus Chatham County, which had previously participated in
116 NCWMN. This latter analysis enabled us to evaluate the comparability of populations that can and cannot
117 contribute to wastewater monitoring in the future.

118 Community demographic and social vulnerability data

119 To summarize population demographics and social vulnerability for sewerred and unsewerred populations,
120 we constructed 23 variables that represent five conceptual domains: demographics, health, housing and
121 transportation, social vulnerability indices, and socio-economic status (SES) (SM Table S1). Most
122 variables clearly fall within one of the five domains, while others straddle multiple domains. We grouped
123 English proficiency within SES because language skills are often related to educational attainment and
124 job prospects. Variables describing race and ethnicity came from the 2020 United States Census
125 Redistricting Data, which were available at the block level geography. Since no other demographic
126 variables were available in the 2020 Census data at the time of our analysis, we also analyzed variables
127 from the 2015-2019 American Community Survey (2015-19 ACS) that captured age, gender, health
128 insurance status, level of education, wealth, English proficiency, housing, employment, and disability
129 status, all of which were available at the tract level geography. Lastly, we downloaded a shapefile with
130 information on the CDC's social vulnerability index (SVI), which were available at the tract level; we
131 summarized the overall SVI percentile rank and the ranks for each of the four SVI themes (socioeconomic
132 status, household composition and disability, minority status and language, and housing type and
133 transportation).

134 Geospatial analysis

135 To prepare the data for a geospatial analysis, we cleaned and joined tabular Census data to TIGER/Line
136 tract or block polygons (US Census Bureau, 2022). We filtered the data to include only the counties in the
137 study area described above. Polygon shapefiles for monitored sewersheds were shared by the NCWMN.
138 For analyses of sewerred versus unsewerred populations, we merged shapefiles of NCWMN monitored

139 sewersheds with shapefiles of WWTPs not monitored by the NCWMN, which we obtained from NC
140 OneMap (North Carolina Department of Information Technology, 2022), to create a single county-level
141 sewerage polygon. We then created polygons that captured the county's unsewered population by
142 removing the sewerage polygon from the county polygon.

143 To summarize the demographics, SES, and SVI of populations residing in the geographies of interest—
144 including monitored individual sewersheds, monitored combined sewersheds, monitored counties,
145 sewerage county areas, unsewered county areas, and the state—we performed spatial intersections, selecting
146 those census blocks or tracts that intersected each polygon of interest. We merged the selected tracts or
147 blocks into a single polygon and calculated summary statistics: percentages that capture the proportion of
148 the total population represented by different demographic groups, the average median household income,
149 and population-weighted averages of SVI ranks. All analyses were performed using either ArcGIS Pro
150 2.9 (ESRI Inc. 2021) or R version 4.1.3 (R Core Team 2022) using the *sf* (Pebesma 2018) and *tidycensus*
151 (Walker and Herman 2022) packages.

152 We designated potentially meaningful differences between populations using a threshold of +/- 5
153 percentage points (pp) for categorical variables (demographic, SES variables, and SVI percentile ranks),
154 and a threshold of +/- 5 percent (%) for continuous variables (median household income). We chose this
155 approach to be conservative and ensure that we did not overlook smaller disparities that were within
156 reported margins of error (MOEs). This is especially relevant for a health-equity focused analysis because
157 smaller groups often have larger MOEs, but a lack of statistical significance should not be interpreted as a
158 lack of meaningful findings. For completeness, we also flagged whether such differences were
159 statistically significant—i.e., whether the differences were within twice of the reported margins of error
160 (MOEs)—for ACS 2015-2019 variables (note: 2020 Census data and SVI data did not include MOEs at
161 the time of this analysis).

162 **Results**

163 Characteristics of WWTPs participating in the NC Wastewater Monitoring Network

164 The WWTPs participating in the NCWMN as of June 2022 covered a broad geographic area of the state
165 and had service populations that ranged in size from 3,500 to 550,000 people and accounted for 1%
166 (Raleigh 3) to 60% (City of Wilson) of the county's population (Table 1). More detailed environmental
167 meta data of the wastewater monitoring program can be found in SM Table S2 and a previous publication
168 (Keshaviah et al. 2023). In three of the 17 counties analyzed, multiple WWTPs were being monitored,
169 which together accounted for 33% (Mecklenburg), 54% (New Hanover), and 75% (Wake) of the
170 respective county's population.

171 Comparison of monitored sewershed with state and county populations

172 As a whole, populations residing in the 25 sewersheds monitored through NCWMN resembled the
173 statewide resident population. Differences between the two groups amounted to less than +/- 5 pp or 5%
174 for 15 of the 23 variables analyzed, including: demographics (percent female, percent African American,
175 percent Asian, percent American Indian/Alaska Native, percent Native Hawaiian or Pacific Islander,
176 percent 65 years and older, percent Hispanic), health status (percent with disability, percent without
177 health insurance), housing and transportation (percent of households without a vehicle, percent group
178 quarters), housing and transportation vulnerability (based on the SVI), and SES (percent below federal
179 poverty line, percent unemployed, percent limited English proficiency) (Figure 1A). However, monitored
180 sewershed populations had fewer White residents (i.e., more minorities), lower overall social vulnerability
181 but higher minority and language vulnerability, a higher number of houses with five or more units, greater
182 educational attainment, a higher median household income and lower socioeconomic vulnerability, and
183 lower vulnerability related to household composition and disability compared to the statewide population
184 (Figure 1B). Among these potentially meaningful differences, only educational attainment reached
185 statistical significance (SM Table S3).

186 When we compared the populations living in the sewersheds monitored by the NCWMN with their
187 respective countywide populations, we found that sewersheds had wide-ranging demographic and social
188 characteristics. Combined monitored sewershed populations generally resembled their countywide

189 population with respect to demographics (percent female, percent over 65 years old, percent Hispanic),
190 health status (percent uninsured, percent disability), and SES (percent limited English speaking, percent
191 below federal poverty, percent unemployed). However, there was a +/- 5 pp or 5% difference between at
192 least one combined monitored sewershed and the county for 12 out of the 23 variables, with the largest
193 observed differences related to race, social vulnerability, median household income, and housing with
194 greater than five units (Figure 2). Monitored sewershed populations had a lower share of Whites
195 compared to countywide populations in 12 of 17 counties (shaded in blue), while African Americans
196 made up a higher share of the monitored sewershed population in 14 of 17 counties, with the largest
197 differences generally occurring in sewersheds in the eastern part of the state (shaded in red). In one
198 county (Jackson County), we observed a potentially meaningful difference in the share of American
199 Indian and Alaska Native residents, which was lower in the monitored sewershed than in the county
200 (note: Jackson County borders the Qualla Boundary, which is home to the sovereign nation of the Eastern
201 Band of the Cherokee Indians). Overall SVI ranks were similar between monitored sewershed populations
202 and countywide populations. However, evaluating the SVI themes individually showed that either or both
203 minority and language vulnerability and housing and transportation vulnerability were higher in most (15
204 out of 17) of monitored sewersheds than countywide (Figure 3). Lastly, the difference in median
205 household income was wide-ranging across sewersheds (-19.8% to +5.8%), with nine sewershed
206 populations having a higher median household than the county and eight having a lower median
207 household income.

208 In the three counties with multiple monitored WWTPs, we noted that populations in the individual
209 monitored sewersheds had differing degrees of similarity to the countywide population. In all three
210 counties, we observed meaningful differences in race, median household income, social vulnerability,
211 educational attainment, and housing with 5 or more units that were not evident when the individual
212 sewersheds were combined and analyzed as a single geographic unit (SM Table S5). For example, in
213 Wake County, the combined sewershed SVI rankings resembled the county SVI rankings even though the

214 six individual sewersheds showed a wide range of SVI rankings across all four SVI themes:
215 socioeconomic status (individual ranged from 0.12-0.51, combined=0.27, county=0.27), household
216 composition and disability (individual=0.16-0.74, combined=0.28, county=0.29), minority status and
217 language (individual=0.44-0.76, combined=0.59, county=0.56), and housing and transportation
218 (individual=0.25-0.63, combined=0.42, county=0.40) (Figure 3). Notably, residents in two Wake County
219 sewersheds, Raleigh and Raleigh 3, appeared to be more disadvantaged than other Wake sewersheds and
220 countywide residents, given their higher social vulnerability overall and across all themes, coupled with a
221 lower educational achievement and lower median household income.

222 Comparison of sewered and unsewered populations

223 When we compared the characteristics of sewered and unsewered populations in nine counties with
224 complete information on WWTP service populations, we found that only four of 23 variables did not
225 meaningfully differ—including percent Asian, percent Native Hawaiian and Pacific Islander, percent
226 female, and percent unemployed—while the remaining 19 variables differed in at least one county (Figure
227 4). Most notably, we found differences in racial and ethnic makeup, median household income, and social
228 vulnerability. In most of the 9 counties, Hispanics and African Americans made up a greater share of the
229 sewered population than the unsewered population (shaded in red). However, the magnitude and direction
230 of effect sometimes varied across counties with a -4.6 to 28.0 pp difference in the share of Hispanics and
231 a 2.4 to 18.0 pp difference in the share of African Americans. We also found that in all but one county
232 (Pitt County), the median household income was lower in the sewered population than the unsewered
233 population, with differences ranging from -30.0% to -0.2%. Contrary to our finding that population
234 residing in NCWMN monitored sewersheds have higher educational attainment than the statewide
235 population, we saw that educational attainment was lower in the sewered population than the unsewered
236 population in 7 of 9 counties (all but Forsyth County and Pitt County), ranging from a -17.0 to -0.2 pp
237 difference. This difference was also statistically significant for Durham and Chatham counties (SM Table
238 S4). Finally, in 7 of 9 counties (all but Jackson County and Pitt County), we found that overall social

239 vulnerability and vulnerability based on each of the four SVI themes were higher among the sewer-
240 ed population than the unsewered population.

241 **Discussion**

242 Our findings indicate that residents of 25 WWTP sewersheds monitored by NCMWN as of June 2022
243 represent the broader North Carolina population fairly well. The share of elderly residents was
244 comparable between monitored sewersheds and state- and county-wide populations, suggesting that
245 wastewater data is capturing infections among groups that experienced more severe COVID-19 disease
246 and higher mortality rates. Further, the share of residents with health insurance, English proficiency, and a
247 vehicle was similar between residents of monitored sewersheds and the respective countywide population,
248 suggesting that there is little bias in the data related to health care access. Given the similarities of
249 demographic and social vulnerability dimensions across most counties, wastewater data collected through
250 the NCWMN can be validly used to describe state and county population health.

251 We did note a few differences that have implications related to health equity. First, we noted that
252 educational attainment was significantly higher among the monitored population than statewide. This
253 finding is important because lower educational attainment has been shown to associate with lower
254 receptivity to public health messaging and higher vaccine hesitancy (Anakpo and Mishi 2022). This may
255 be explained by the fact that the initial sites chosen for monitoring were located near the university
256 laboratories that piloted wastewater testing. Second, in the three counties with multiple monitored
257 sewersheds, differences between characteristics of residents in individual monitored sewersheds versus
258 countywide suggest that, had NCWMN monitored only one of the sewersheds in the county, or if the site
259 composition in these counties changes over time, the wastewater data may not be representative of the
260 larger population. Third, we found that sewer- ed populations have significantly lower educational
261 attainment, a lower share of elderly residents, and a higher share living below the poverty level compared
262 to unsewered residents. In other words, sewer- ed populations may be more vulnerable than unsewered
263 populations, suggesting that wastewater data is likely to capture the health status of vulnerable residents.

264 Our analyses also highlight that minority populations may be over-represented in the state’s wastewater
265 data. African Americans represented a higher share of monitored sewershed residents than countywide,
266 while Whites often comprised a smaller share of monitored sewershed residents than county- or state-
267 wide. Also, vulnerability related to minority status and language was greater in monitored sewersheds
268 than statewide. Together, these findings have two implications. First, they suggest that, by better
269 representing potentially vulnerable racial and ethnic minorities, wastewater monitoring may have filled
270 critical gaps in clinical case data, which likely underrepresented Black and Hispanic communities early in
271 the pandemic. More recent research has also found that, in the summer of 2022 (when the Omicron
272 variant was dominant), COVID-19 cases were severely underestimated, with differences particularly
273 pronounced among Black and Hispanic populations, younger adults ages 18 to 24, and those with lower
274 income and less education (Qasmieh et al. 2023). The second implication of having a higher share of
275 minority residents in monitored sewersheds versus county- or state-wide is that it creates a risk of
276 inaccurate health messaging to the public. Given that racial and ethnic minorities have seen higher SARS-
277 CoV-2 infection rates than White, non-Hispanic populations (CDC 2023; Mackey et al. 2021), wastewater
278 data that overrepresents these groups could lead to inflated COVID-19 infection estimates.

279 Although we found that populations monitored by NCWMN were generally representative of the larger
280 county- and state-wide population, there were some limitations to our analysis. First, our findings may not
281 hold as monitoring sites are added or removed in the future. At the time of our analysis, which was based
282 on a snapshot as of June 2022, the NCWMN was monitoring 25 sewersheds in 17 counties. Roughly one
283 year later— as of June 14, 2023—the NCWMN monitors 50 sites in 32 counties. Accordingly, continued
284 efforts will be needed to uncover any bias that emerges in the wastewater data. Expanding statewide
285 monitoring to include community onsite wastewater treatment systems could also strengthen equitable
286 coverage. This strategy may be particularly useful in NC, since roughly 50% of state residents use septic
287 systems (NCDHHS 2022), and we found that NC’s sewerred and unsewerred populations differ with
288 respect to demographics and social vulnerability.

289 A second limitation is that our geospatial analysis may have misclassified some residents as being part of
290 the monitored sewershed population when they were not. To aggregate data to the sewershed level, we
291 utilized a spatial intersect to select tracts or blocks that intersected the sewershed polygon. This selection
292 method ensured that the selection for small or narrow sewersheds was greater than zero, but since neither
293 geography perfectly aligns with the sewershed boundary, it possibly over-estimated the number of
294 persons in the sewershed when tracts or blocks partially extend outside the sewershed boundary. Future
295 studies using hi-resolution gridded population data would more accurately assign populations to the
296 sewershed (Depsky et al. 2022). We also assumed that all homes inside the sewershed boundary are
297 connected to the sewer, even though some may use onsite septic systems. We were unable to discern the
298 magnitude or direction of the resulting bias from this assumption because septic system location data are
299 not readily available for the state.

300 A third limitation is that our analysis could not examine representativeness related to SARS-CoV-2 fecal
301 shedding rates. People that shed little or no virus in their feces will not be represented in wastewater data,
302 and preliminary research suggests that demographic and geographic features may influence viral shedding
303 rates. For example, early in the pandemic, Parasa et al. (2020) found that fecal shedding rates varied
304 substantially across eight studies included in their meta-analysis, estimating that, on average, 41% of
305 confirmed COVID-19 cases (range=17% to 80%) shed the virus in their stools. More recently, Prasek et
306 al. (2023) noted differences in estimated shedding rates across communities of differing ages, ethnicity,
307 and socioeconomic composition, as well as over time, as the dominant variant changed (though it is worth
308 noting that these findings were subject to ecological fallacy and lacked the use of multivariate regression
309 modeling to control for confounding factors).

310 Lastly, it is important to interpret our findings in the context of known limitations and biases in the
311 underlying Census data. Data on race and ethnicity collected during the 2020 US Census were subjected
312 to a new disclosure avoidance system called differential privacy, which added statistical noise to the
313 published data products to shield sensitive information from discovery. However, the amount of statistical

314 noise added to the data is not constant and small demographic groups and geographic areas were infused
315 with more noise to reduce the risk of re-identification (Garfinkel 2022). We aggregated Census block data
316 to larger geographies, which should minimize inaccuracies associated with differential privacy. Moreover,
317 the Demographic Analysis, one of the leading indicators of data quality for decennial censuses, showed a
318 record undercount of Hispanics during the 2020 Census (Cohn and Passel 2022). Although we did not
319 discover a meaningful difference in the percent of the population that was Hispanic between monitored
320 sewershed and the county or the state, it is possible that a difference was obscured by undercounting.

321 **Conclusions**

322 Evidence-based public health decisions need to be informed by complete, high-quality data that represent
323 the community. Our analyses confirm that wastewater data from across North Carolina well represents
324 county- and state-wide populations, and that by capturing the health information of vulnerable
325 populations better than clinical data—particularly when clinical resources are in short supply—
326 wastewater data can promote health equity. The in-depth geospatial analyses we conducted here identify
327 underlying bias in wastewater data, and in doing so, can also help officials recognize when and how to
328 tailor public health messaging and response accordingly.

329 **Acknowledgements**

330 The authors would like to acknowledge the contributions of several partners to this work, including
331 Virginia Guidry, Ariel Christensen, and Steven Berkowitz from the North Carolina Department of Health
332 and Human Services. Financial support for this research was provided by the authors' institutions—
333 Mathematica, NCDHHS, and North Carolina State University.

334

335 **Data Sharing**

336 The data that support the findings of this study are available from the U.S. Census Data API
337 <https://www.census.gov/data/developers/data-sets/census-microdata-api.html>. Shapefiles of the WWTP

338 sewer service areas (not including those provided by NCWMN) are available from NC OneMap
339 <https://www.nconemap.gov/datasets/nconemap::type-a-current-public-sewer-systems-2004/about> .

340 **References**

- 341 Anakpo G, Mishi S. 2022. Hesitancy of COVID-19 vaccines: Rapid systematic review of the
342 measurement, predictors, and preventive strategies. *Human Vaccines & Immunotherapeutics*
343 18:2074716; doi:10.1080/21645515.2022.2074716.
- 344 CDC. 2023. Cases, Data, and Surveillance. Centers for Disease Control and Prevention. Available:
345 [https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-](https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html)
346 [death-by-race-ethnicity.html](https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html) [accessed 26 April 2023].
- 347 Cohn D, Passel JS. 2022. Key facts about the quality of the 2020 census. Pew Research Center.
- 348 Depsky NJ, Cushing L, Morello-Frosch R. 2022. High-resolution gridded estimates of population
349 sociodemographics from the 2020 census in California. *PLOS ONE* 17:e0270746;
350 doi:10.1371/journal.pone.0270746.
- 351 ESRI Inc. 2021. ArcGIS Pro (Version 2.9).
- 352 Garfinkel S. 2022. Differential Privacy and the 2020 US Census · Winter 2022. Available: [https://mit-](https://mitserc.pubpub.org/pub/differential-privacy-2020-us-census/release/1)
353 [serc.pubpub.org/pub/differential-privacy-2020-us-census/release/1](https://mitserc.pubpub.org/pub/differential-privacy-2020-us-census/release/1) [accessed 26 April 2023].
- 354 González-Mariño I, Baz-Lomba JA, Alygizakis NA, Andrés-Costa MJ, Bade R, Bannwarth A, et al.
355 2020. Spatio-temporal assessment of illicit drug use at large scale: evidence from 7 years of
356 international wastewater monitoring. *Addiction* 115:109–120; doi:10.1111/add.14767.
- 357 Hart OE, Halden RU. 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by
358 wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and
359 challenges. *Science of The Total Environment* 730:138875; doi:10.1016/j.scitotenv.2020.138875.
- 360 Hovi T, Shulman LM, van der Avoort H, Deshpande J, Roivainen M, DE Gourville EM. 2012. Role of
361 environmental poliovirus surveillance in global polio eradication and beyond. *Epidemiol Infect*
362 140:1–13; doi:10.1017/S095026881000316X.
- 363 Keshaviah A. 2021. Next steps for wastewater testing to help end this pandemic — and prevent the next
364 one. *STAT*. Available at [https://www.statnews.com/2021/06/24/wastewater-testing-](https://www.statnews.com/2021/06/24/wastewater-testing-infrastructure-improvements-public-health/)
365 [infrastructure-improvements-public-health/](https://www.statnews.com/2021/06/24/wastewater-testing-infrastructure-improvements-public-health/)
- 366 Keshaviah A, Huff I, Hu XC, Guidry V, Christensen A, Berkowitz S, et al. 2023. Separating signal from
367 noise in wastewater data: An algorithm to identify community-level COVID-19 surges in real
368 time. *Proceedings of the National Academy of Sciences* 120:e2216021120;
369 doi:10.1073/pnas.2216021120.
- 370 Kim SR. 2020. Which Cities Have The Biggest Racial Gaps In COVID-19 Testing Access?
371 *FiveThirtyEight*.
- 372 Mackey K, Ayers CK, Kondo KK, Saha S, Advani SM, Young S, et al. 2021. Racial and Ethnic
373 Disparities in COVID-19–Related Infections, Hospitalizations, and Deaths. *Ann Intern Med*
374 174:362–373; doi:10.7326/M20-6306.
- 375 National Academies Sciences Engineering and Medicine,. 2023. Wastewater-based Disease Surveillance
376 for Public Health Action |The National Academies Press. Available:

- 377 [https://nap.nationalacademies.org/catalog/26767/wastewater-based-disease-surveillance-for-](https://nap.nationalacademies.org/catalog/26767/wastewater-based-disease-surveillance-for-public-health-action)
378 [public-health-action](https://nap.nationalacademies.org/catalog/26767/wastewater-based-disease-surveillance-for-public-health-action) [accessed 7 April 2023].
- 379 Naughton CC, Roman FA, Alvarado AGF, Tariqi AQ, Deeming MA, Bibby K, et al. 2021. Show us the
380 Data: Global COVID-19 Wastewater Monitoring Efforts, Equity, and Gaps.
381 2021.03.14.21253564; doi:10.1101/2021.03.14.21253564.
- 382 NCDHHS. 2022. EHS: On-Site Wastewater Treatment and Dispersal Systems Program Resources.
383 Available: <https://ehs.dph.ncdhhs.gov/oswp/resources.htm> [accessed 26 April 2023].
- 384 NCDHHS. 2021. NCDHHS Doubles Wastewater Monitoring Sites to Track the Spread of COVID-19.
385 Available: [https://www.ncdhhs.gov/news/press-releases/2021/07/22/ncdhhs-doubles-wastewater-](https://www.ncdhhs.gov/news/press-releases/2021/07/22/ncdhhs-doubles-wastewater-monitoring-sites-track-spread-covid-19)
386 [monitoring-sites-track-spread-covid-19](https://www.ncdhhs.gov/news/press-releases/2021/07/22/ncdhhs-doubles-wastewater-monitoring-sites-track-spread-covid-19) [accessed 27 January 2023].
- 387 Noble R, de los Reyes F, Harris A, Stewart J, Cahoon L, Engel L, et al. 2021. Tracking SARS-CoV-2 in
388 the Wastewater Across a Range of North Carolina Municipalities.
- 389 Noh JY, Yoon JG, Seong H, Choi WS, Sohn JW, Cheong HJ, et al. 2020. Asymptomatic infection and
390 atypical manifestations of COVID-19: Comparison of viral shedding duration. *Journal of*
391 *Infection* 81:816–846; doi:10.1016/j.jinf.2020.05.035.
- 392 North Carolina Department of Information Technology,. 2022. NC OneMap. Available:
393 <https://www.nconemap.gov/> [accessed 7 April 2023].
- 394 Olesen SW, Young C, Duvallet C. 2022. The Effect of Septic Systems on Wastewater-Based
395 Epidemiology.
- 396 Parasa S, Desai M, Thoguluva Chandrasekar V, Patel HK, Kennedy KF, Roesch T, et al. 2020.
397 Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients With Coronavirus
398 Disease 2019: A Systematic Review and Meta-analysis. *JAMA Network Open* 3:e2011335;
399 doi:10.1001/jamanetworkopen.2020.11335.
- 400 Pebesma E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*
401 10:439–446; doi:10.32614/RJ-2018-009.
- 402 Prasek SM, Pepper IL, Innes GK, Slinski S, Betancourt WQ, Foster AR, et al. 2023. Variant-specific
403 SARS-CoV-2 shedding rates in wastewater. *Science of The Total Environment* 857:159165;
404 doi:10.1016/j.scitotenv.2022.159165.
- 405 Qasmieh SA, Robertson MM, Teasdale CA, Kulkarni SG, Jones HE, McNairy M, et al. 2023. The
406 prevalence of SARS-CoV-2 infection and long COVID in U.S. adults during the BA.4/BA.5
407 surge, June–July 2022. *Preventive Medicine* 169:107461; doi:10.1016/j.jypmed.2023.107461.
- 408 R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for
409 Statistical Computing:Vienna, Austria.
- 410 Rader B, Astley CM, Sy KTL, Sewalk K, Hswen Y, Brownstein JS, et al. 2020. Geographic access to
411 United States SARS-CoV-2 testing sites highlights healthcare disparities and may bias
412 transmission estimates. *J Travel Med* 27:taaa076; doi:10.1093/jtm/taaa076.

- 413 US Census Bureau,. 2022. TIGER/Line Shapefiles. Census.gov. Available:
414 <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html> [accessed
415 7 April 2023].
- 416 US EPA Office of Water,. 2015. Learn about Small Wastewater Systems. Available:
417 <https://www.epa.gov/small-and-rural-wastewater-systems/learn-about-small-wastewater-systems>
418 [accessed 27 January 2023].
- 419 Walker K, Herman M. 2022. *tidycensus: Load US Census Boundary and Attribute Data as “tidyverse”*
420 *and ‘sf’-Ready Data Frames*.
- 421 World Bank. n.d. World Bank Country and Lending Groups – World Bank Data Help Desk. Available:
422 [https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups)
423 [lending-groups](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups) [accessed 7 December 2022].
- 424 Zhou R, Li F, Chen F, Liu H, Zheng J, Lei C, et al. 2020. Viral dynamics in asymptomatic patients with
425 COVID-19. *International Journal of Infectious Diseases* 96:288–290;
426 doi:10.1016/j.ijid.2020.05.030.
- 427
- 428

429 **Tables**

430 **Table 1. North Carolina sites monitored by the NCWMN as of June 2022**

Site name	County name	Catchment population size	County population size	% of the county population served
Laurinburg	Scotland	15,527	34,823	45%
Tuckaseegee	Jackson ^a	13,296	43,109	31%
Marion	McDowell	8,459	45,756	18%
Beaufort	Carteret ^a	3,500	69,473	5%
Roanoke Rapids	Halifax	14,320	69,493	21%
City of Wilson	Wilson	49,384	81,801	60%
Chapel Hill – Carrboro	Orange	78,141	148,476	53%
Greenville	Pitt ^a	89,616	180,742	50%
Wilmington City	New Hanover ^a	58,361	234,473	25%
New Hanover County (North)	New Hanover ^a	67,743	234,473	29%
South Durham	Durham ^a	108,105	321,488	34%
Fayetteville -Rockfish Creek	Cumberland	151,589	335,509	45%
MSD of Buncombe County	Buncombe	173,000	378,608	46%
Winston Salem - Salem	Forsyth ^a	178,000	382,295	47%
Jacksonville	Onslow	41,819	204,576	20%
Greensboro, North Buffalo	Guilford	135,821	537,174	25%
Charlotte 1	Mecklenburg ^a	68,685	1,110,356	6%
Charlotte 2	Mecklenburg ^a	182,501	1,110,356	16%
Charlotte 3	Mecklenburg ^a	120,000	1,110,356	11%
Raleigh	Wake ^a	550,000	1,111,761	49%
Raleigh 2	Wake ^a	37,020	1,111,761	3%
Raleigh 3	Wake ^a	7,648	1,111,761	1%
Cary 1	Wake ^a	84,189	1,111,761	8%
Cary 2	Wake ^a	74,331	1,111,761	7%
Cary 3	Wake ^a	75,886	1,111,761	7%

431 Note: Sites are listed in order of ascending county population size. MSD=Metropolitan Sewerage District; WWTP=wastewater
 432 treatment plant.

433 ^a Indicates counties included in the sewered vs unsewered county analysis. Chatham county was not actively being monitored in
 434 June 2022 so it is not included here, but it is included in the sewered versus non-sewered analysis.

435

436 **Figure captions**

437 **Figure 1.** Demographic characteristics between the population in monitored sewersheds and the statewide
438 population: A) difference between the population covered by monitored sewersheds and the statewide
439 population is smaller than 5 percentage points; B) the difference between the population covered by
440 monitored sewersheds and the statewide population equals or is greater than 5 percentage points.

441 Note: The error bar represents the 95% confidence interval. 95% confidence interval which was only
442 calculated for variables in the ACS 2015-2019 data because the MOE information wasn't available for
443 2020 census at the time of the analysis. SVI is not a percent of the total population but a percentile rank.
444 Only one variable (% Bachelor's degree or higher) has a statistically significant difference (the difference
445 is greater than twice the margin of error, shown with an asterisk (*)).

446 **Figure 2.** Demographic differences between the population in NCWMN monitored sewersheds and the
447 respective counties.

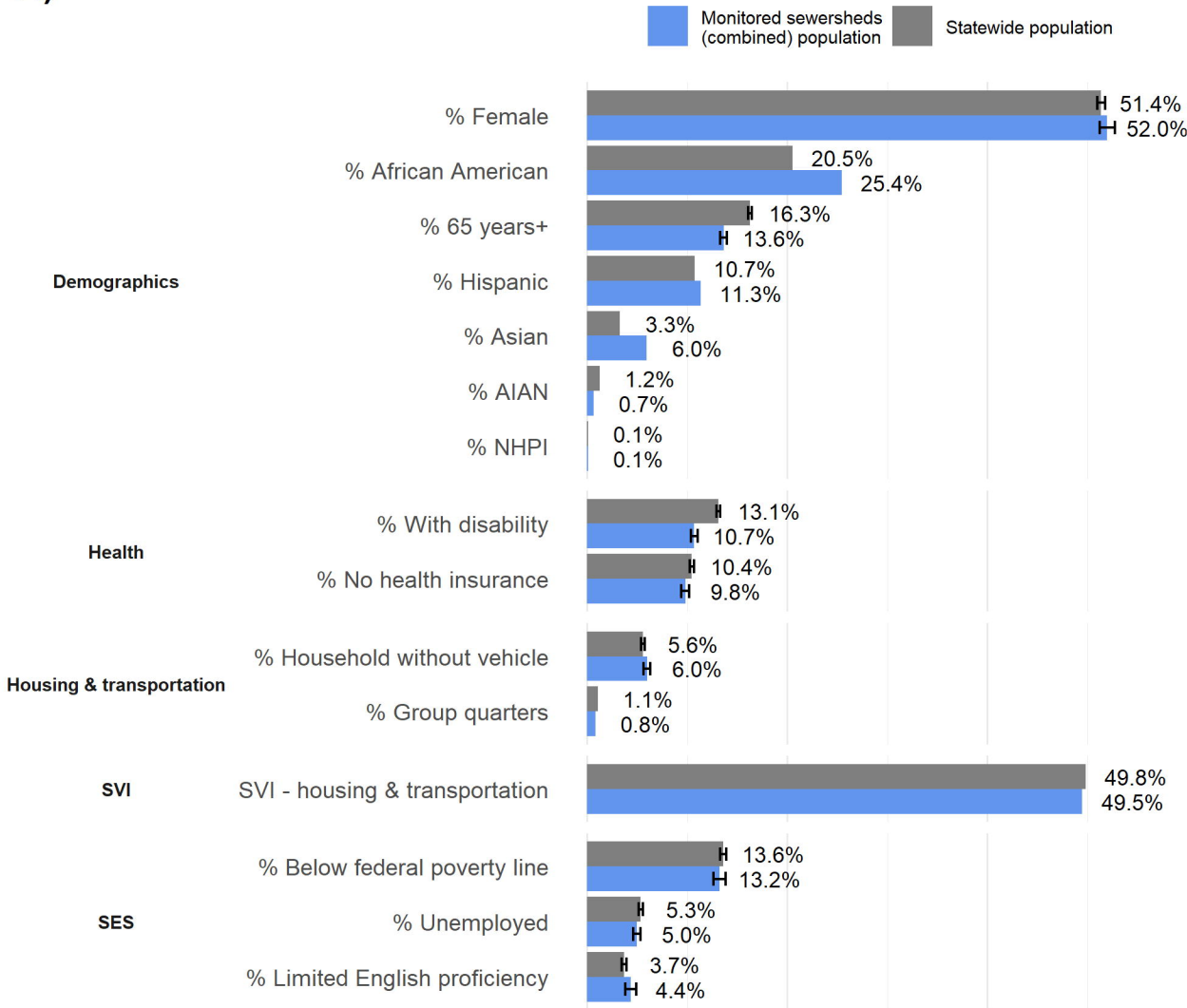
448 Note: Only demographic variables with more than a 5% difference (monitored – county) are included. For
449 percentage variables, the difference is calculated by the monitored population minus the county
450 population (i.e. percentage point difference). For income, the difference is calculated as the percent
451 difference (i.e. income for the monitored population minus income for the county population, then
452 divided by the income for the county population). Counties are ranked from west to east based on the
453 location of the county centroids. Blocks highlighted with black outline are both meaningfully different
454 and statistically significantly different (difference greater than twice the margin of error).

455 **Figure 3.** Social vulnerability percentile rankings in individual monitored NC sewersheds and respective
456 counties. Maps show the four SVI themes: (a) socioeconomic status, (b) household composition and
457 disability, (c) minority and language, and (d) housing and transportation.

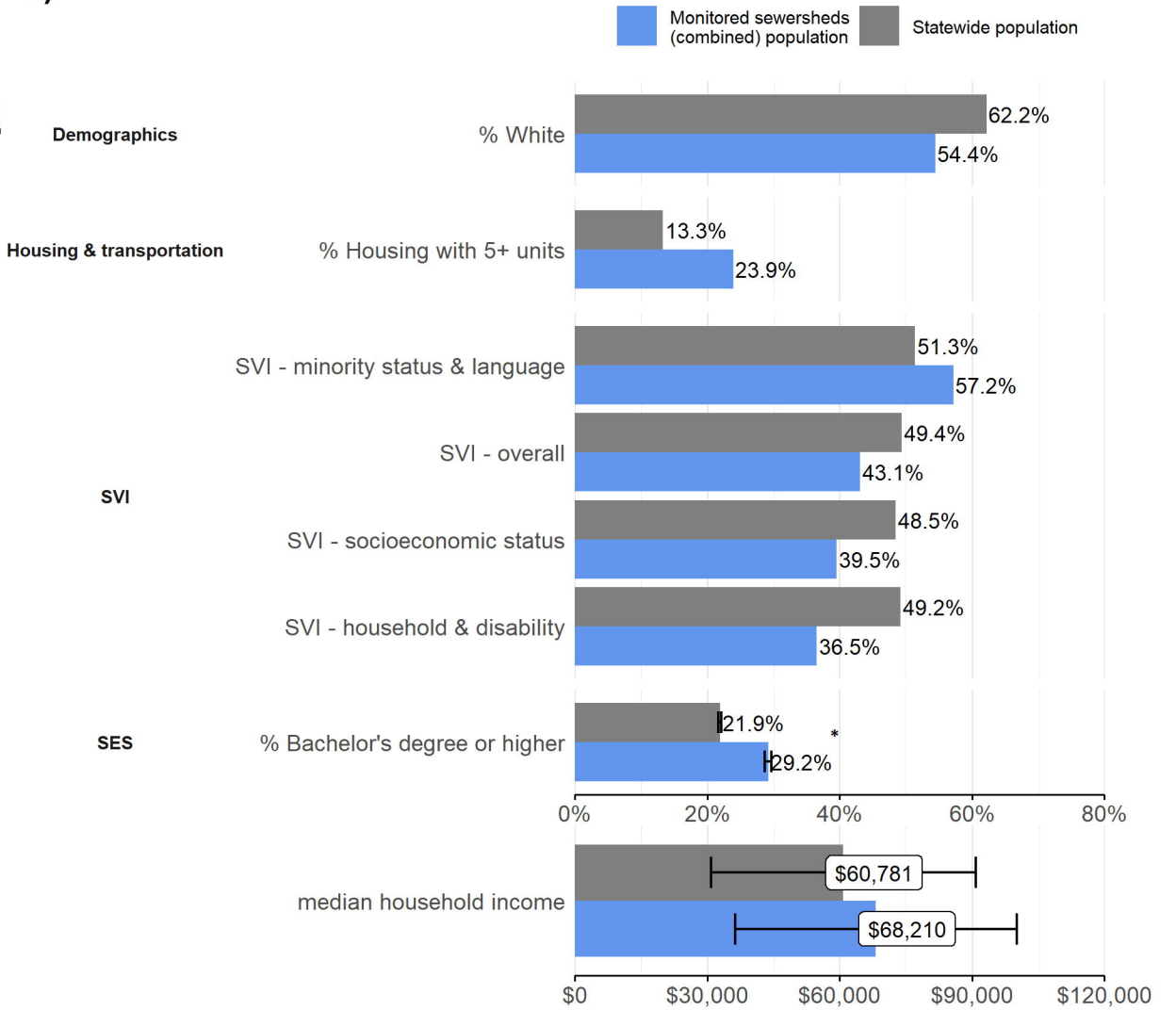
458 **Figure 4.** Demographic differences between the sewered population and the unsewered population.

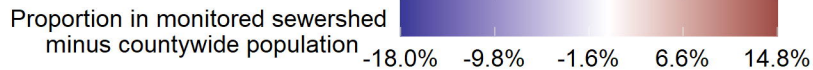
459 Note: Only demographic variables with more than a 5% difference (sewered – unsewered) are included.
460 For percentage variables, the difference is calculated by the sewered population minus the unsewered
461 population (i.e. percentage point difference). For income, the difference is calculated as the percent
462 difference (i.e. income for the sewered population minus income for the unsewered population, then
463 divided by the income for the unsewered population). Counties are ranked from west to east based on the
464 location of the county centroids. Blocks highlighted with black outline are both meaningfully different
465 and statistically significantly different (difference greater than twice the margin of error).

A)

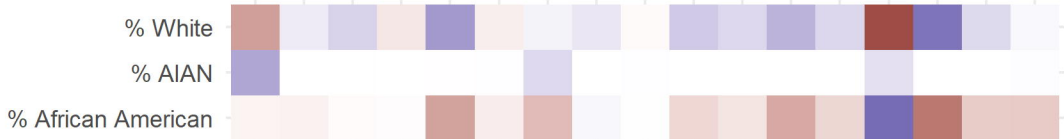


B)

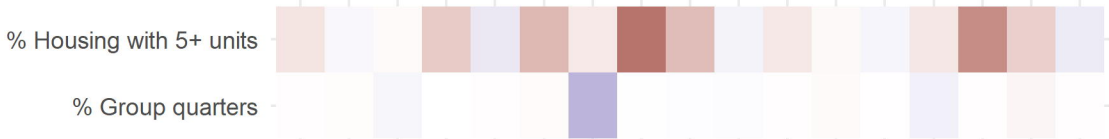




Demographics



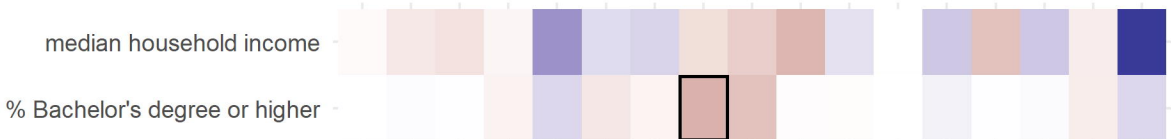
Housing & transportation



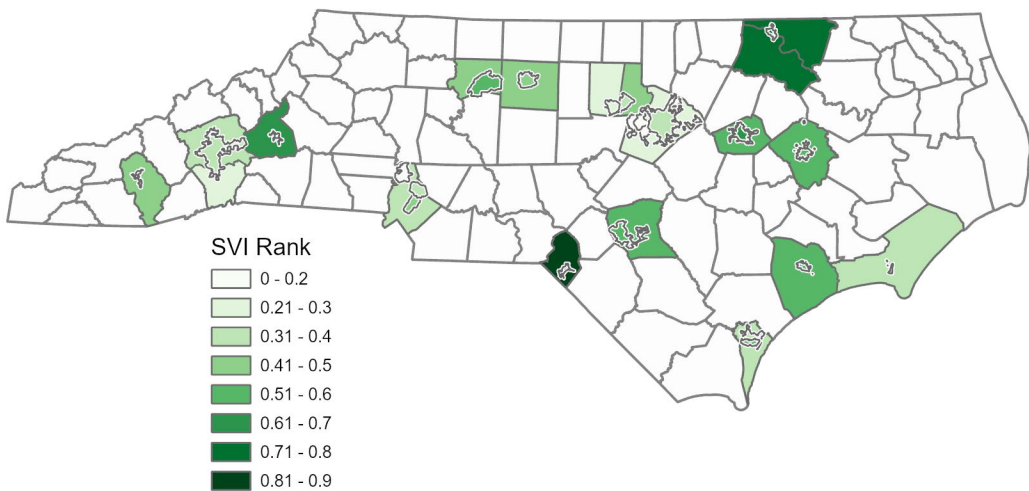
SVI



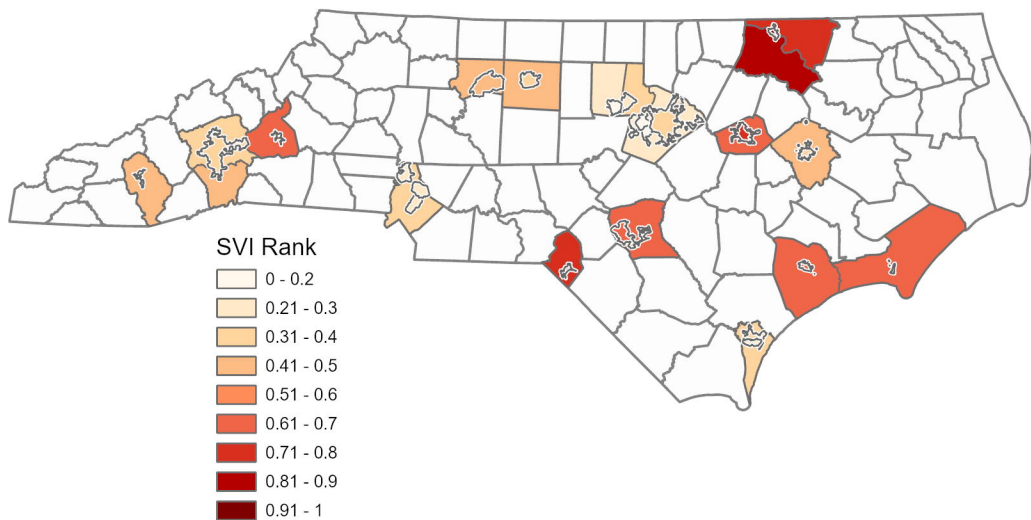
SES



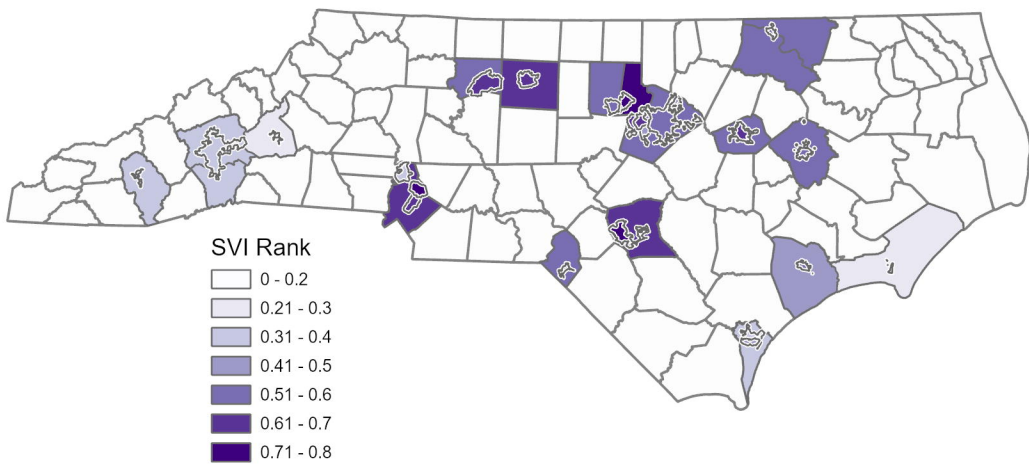
(a)



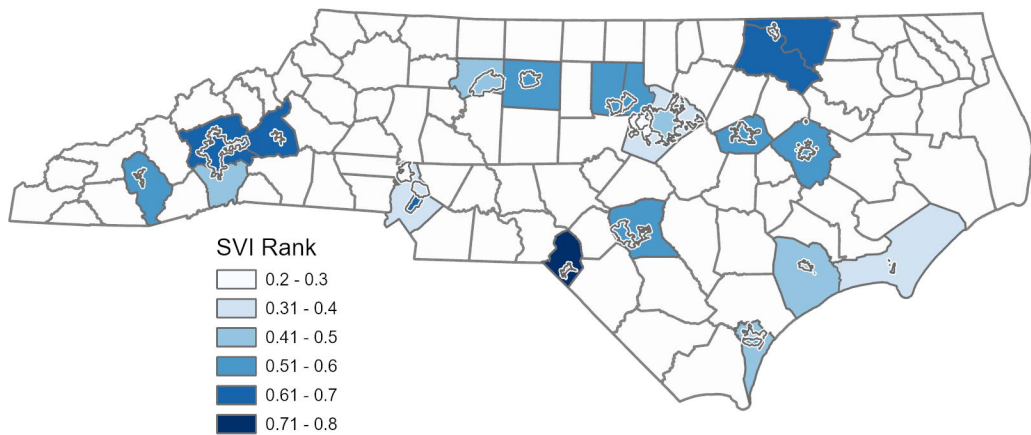
(b)



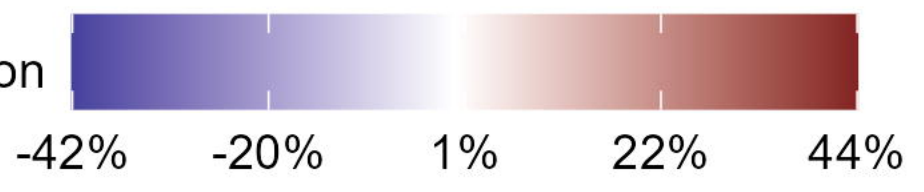
(c)



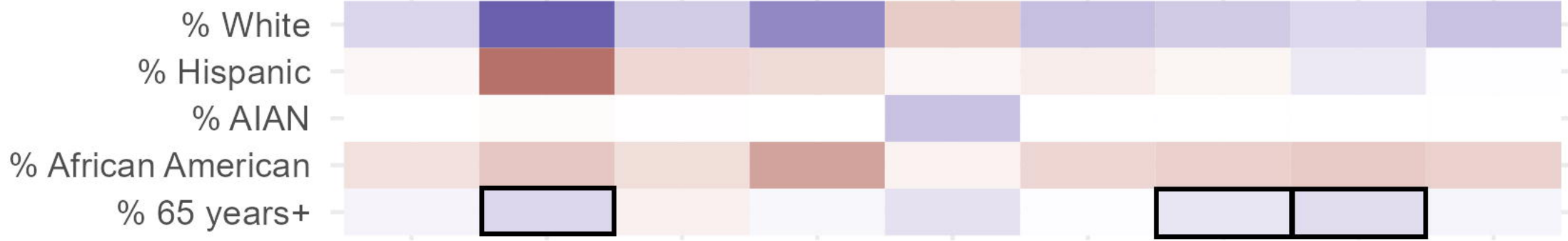
(d)



Proportion in sewerred population
minus unsewerred population



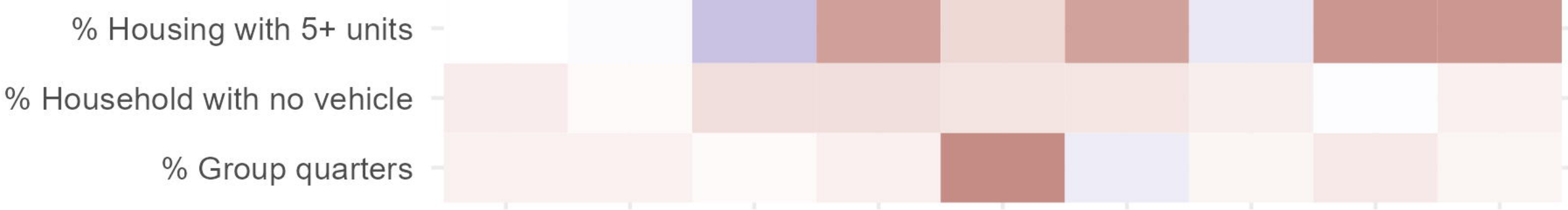
Demographics



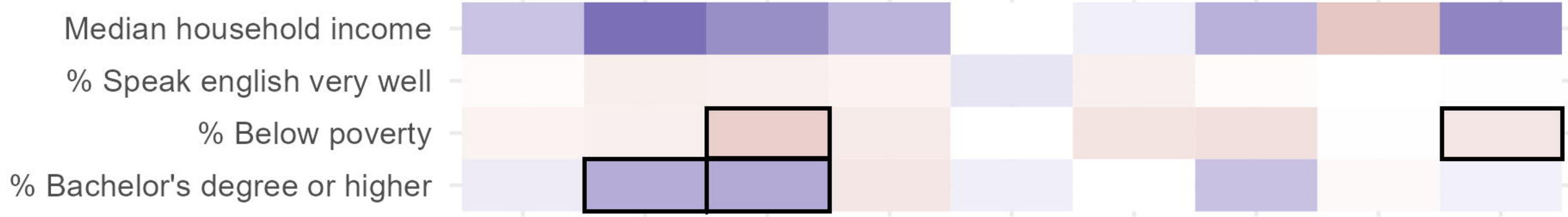
Health



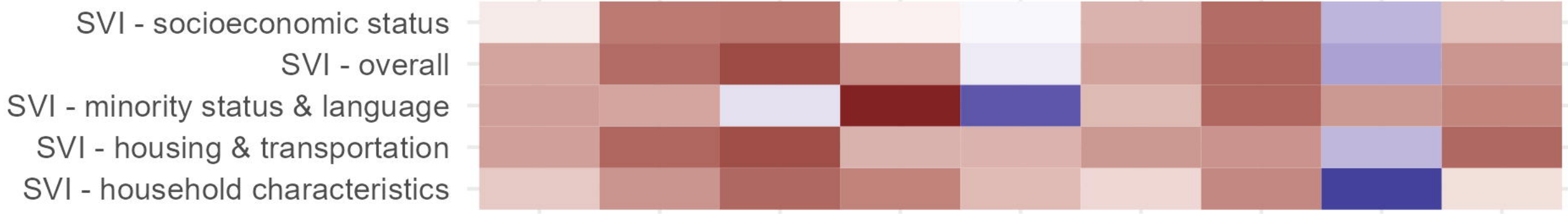
Housing & transportation



SES



SVI



Carteret County
Chatham County
Durham County
Forsyth County
Jackson County
Mecklenburg County
New Hanover County
Pitt County
Wake County

medRxiv preprint doi: <https://doi.org/10.1101/2023.10.06.23296680>; this version posted October 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.