Assessing the Impacts of Differential Privacy on Public Health Surveillance at Varying Geographic Resolutions

Angie Werner, PhD, MPH; Nicholas Skaff, PhD; Adam Vaughan, PhD, MPH

October 24, 2023

2023 FCSM Research & Policy Conference

Q: What is differential privacy?

- Creates uncertainty about the identities of people behind the numbers and allows for measurement of privacy risk.
- Adds precisely calibrated “statistical noise” to data tables published by Census (inserts small differences into counts of people and households) to protect each respondent’s identity.

- Invariants (reported as enumerated)
  - Total population (state and state-equivalents level)
  - Total housing units (census block level) – not block-level resident population
  - Number of group quarters by facilities type (census block level) – not population in group quarters

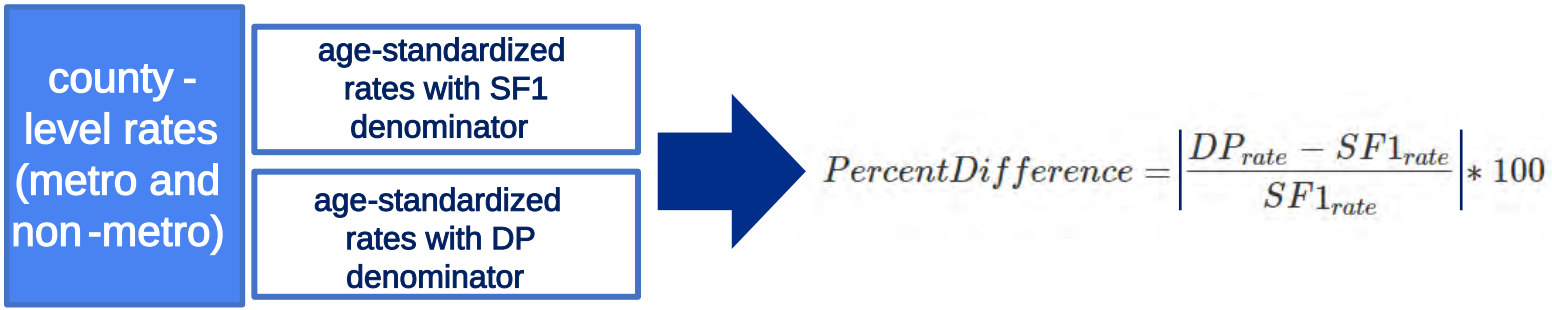
County level: Assessing the impact on age-adjusted incidence of COVID-19

Data and Methods

Data

- Numerator: Number of COVID-19 cases in 2020 by county, race, and age*^
- Denominators: Summary File 1 (SF1) 2010 population data and differential privacy (DP) demonstration data (v4-3-2023)

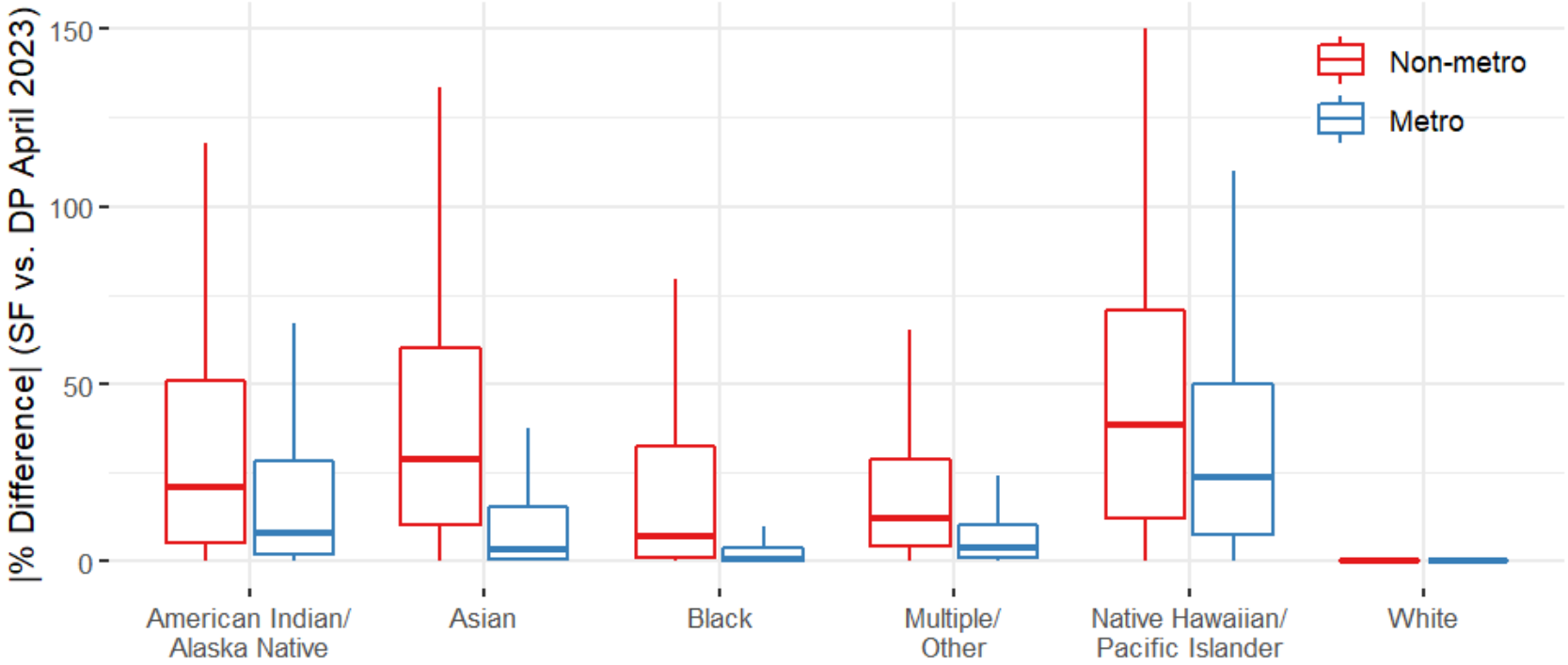
Methods



*included if case report date, age, race, and county of residence were submitted

^limitation for all use cases: 2010 published data used swapping instead of disclosure avoidance

County-level age-adjusted incidence rates for people from racial minority groups were disproportionately affected.



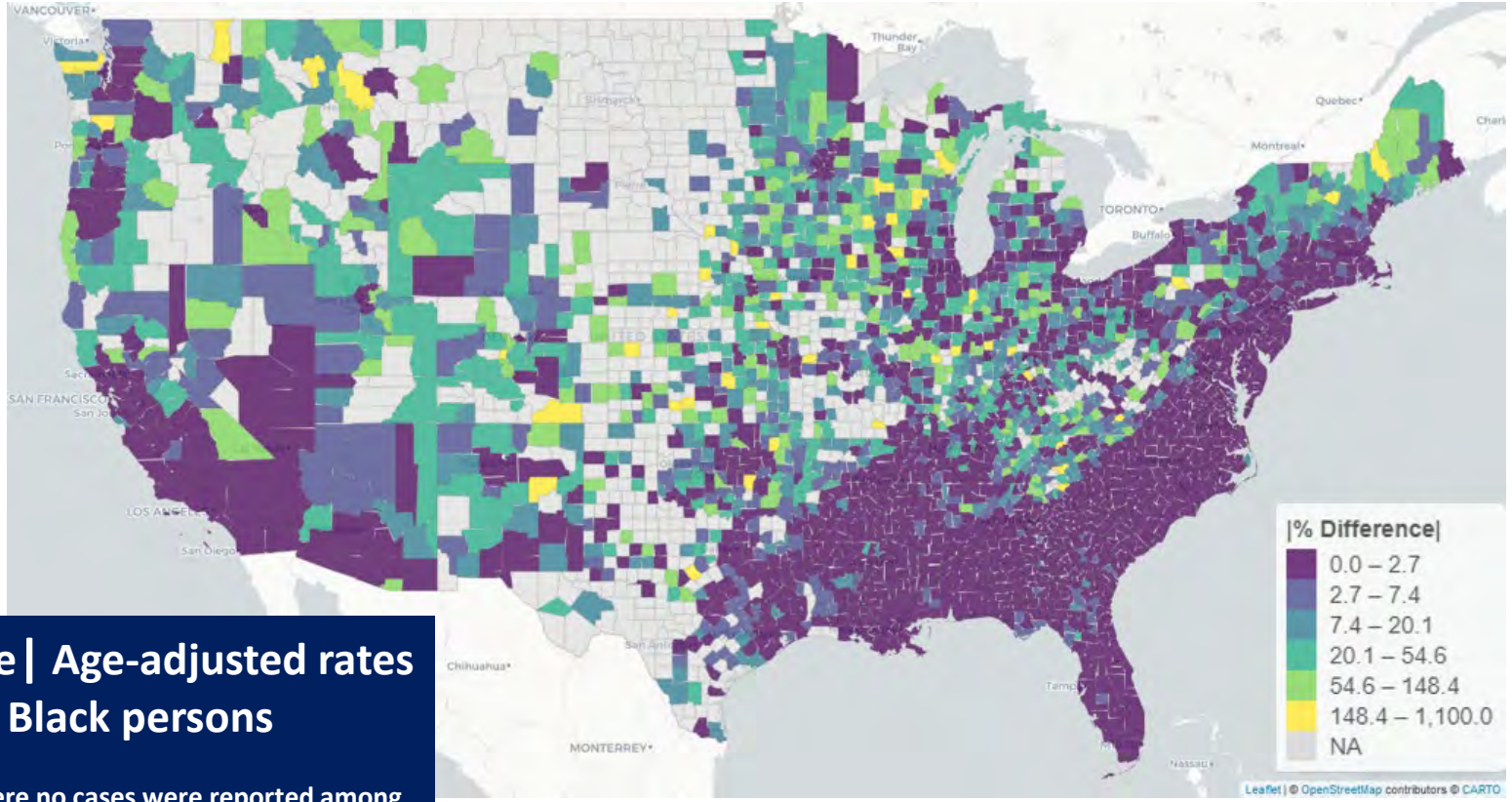
Source: US Census Bureau 2023; CDC Environmental Public Health Tracking Program

COVID-19

Age-adjusted incidence rates for people from racial minority groups were disproportionately affected.

Race/Metro or Non-Metro	Median %	Maximum %
AI/AN – Non-metro	20.8	700
AI/AN – Metro	8.2	700
Asian – Non-metro	28.6	700
Asian – Metro	3.2	559
Black – Non-metro	7.0	900
Black – Metro	0.7	417
Multiple/Other – Non-metro	12.3	1,523.1
Multiple/Other – Metro	3.7	226.1
Native Hawaiian/Other Pacific Islander – Non-metro	38.4	772.9
Native Hawaiian/Other Pacific Islander – Metro	23.8	743.4
White – Non-metro	0.2	52.7
White - Metro	0.04	6.9

COVID-19



|% difference| Age-adjusted rates SF vs. DP for Black persons

NA values occur where no cases were reported among Black persons in that county during 2020.

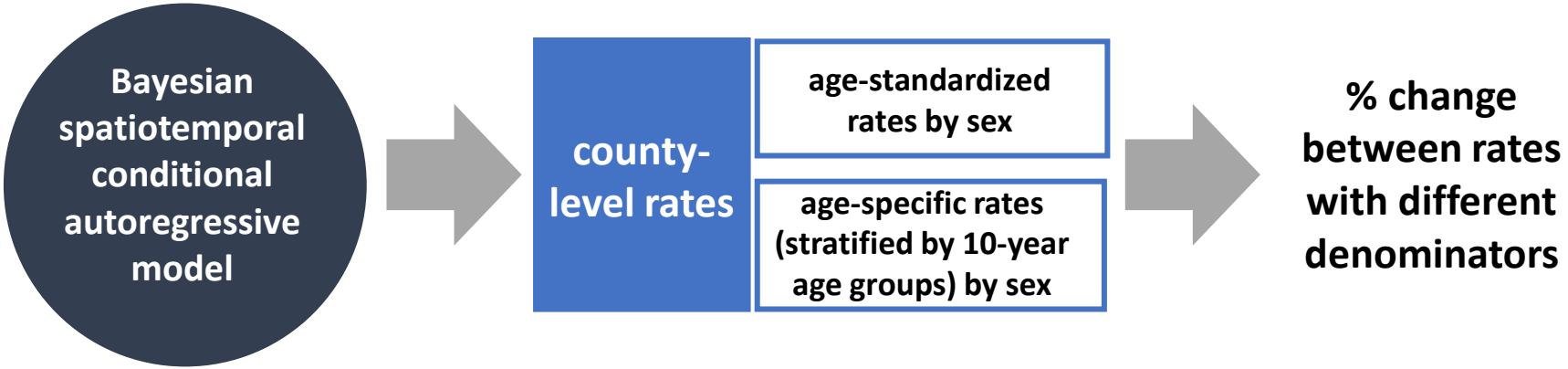
County level: Assessing the impact on heart disease mortality overall and by sub-group

Data and Methods

Data

- Numerator: 2010 county-level heart disease death counts from National Vital Statistics System
- Denominators: National Center for Health Statistics bridged-race estimates and DP demonstration data (v4-3-2023)

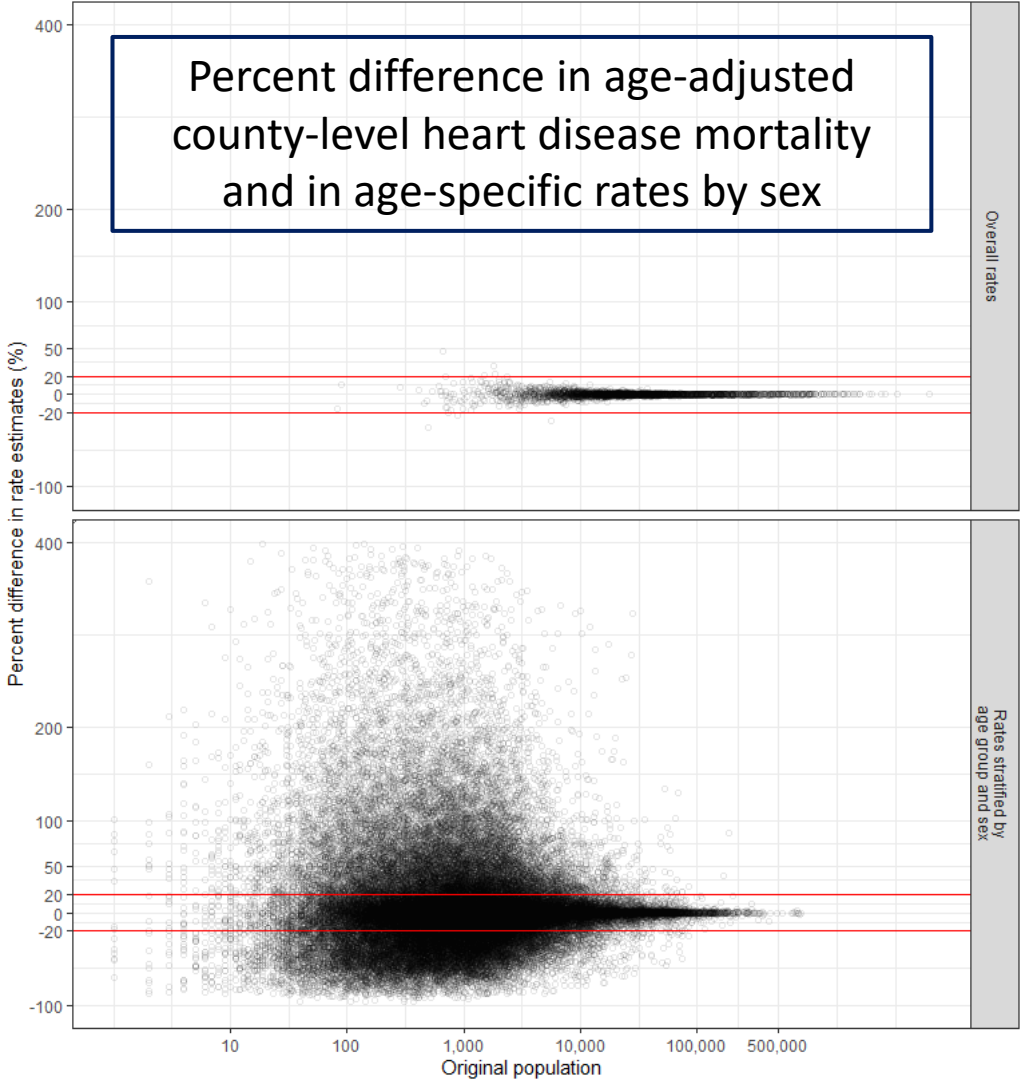
Methods



Heart Disease Mortality

Positive differences indicate the rate estimated using DP was a higher value.

*Outliers truncated.



County and census tract: Assessing the impact on age-adjusted rates of asthma emergency department (ED) visits and acute myocardial infarction (AMI) hospitalizations

Hosp/ED

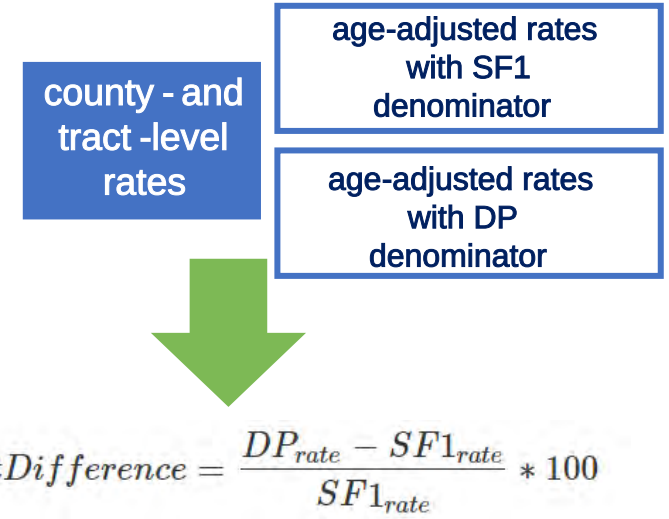
Data and Methods

Data

- **County-level asthma**
  - ED visit counts from 30 states
  - AMI hospitalization counts from 31 states
- **Census tract-level asthma**
  - ED visit counts from 6 states
  - AMI hospitalization counts from 7 states
- **Denominators:** SF1 2010 population data and DP demonstration data (v4-3-2023)

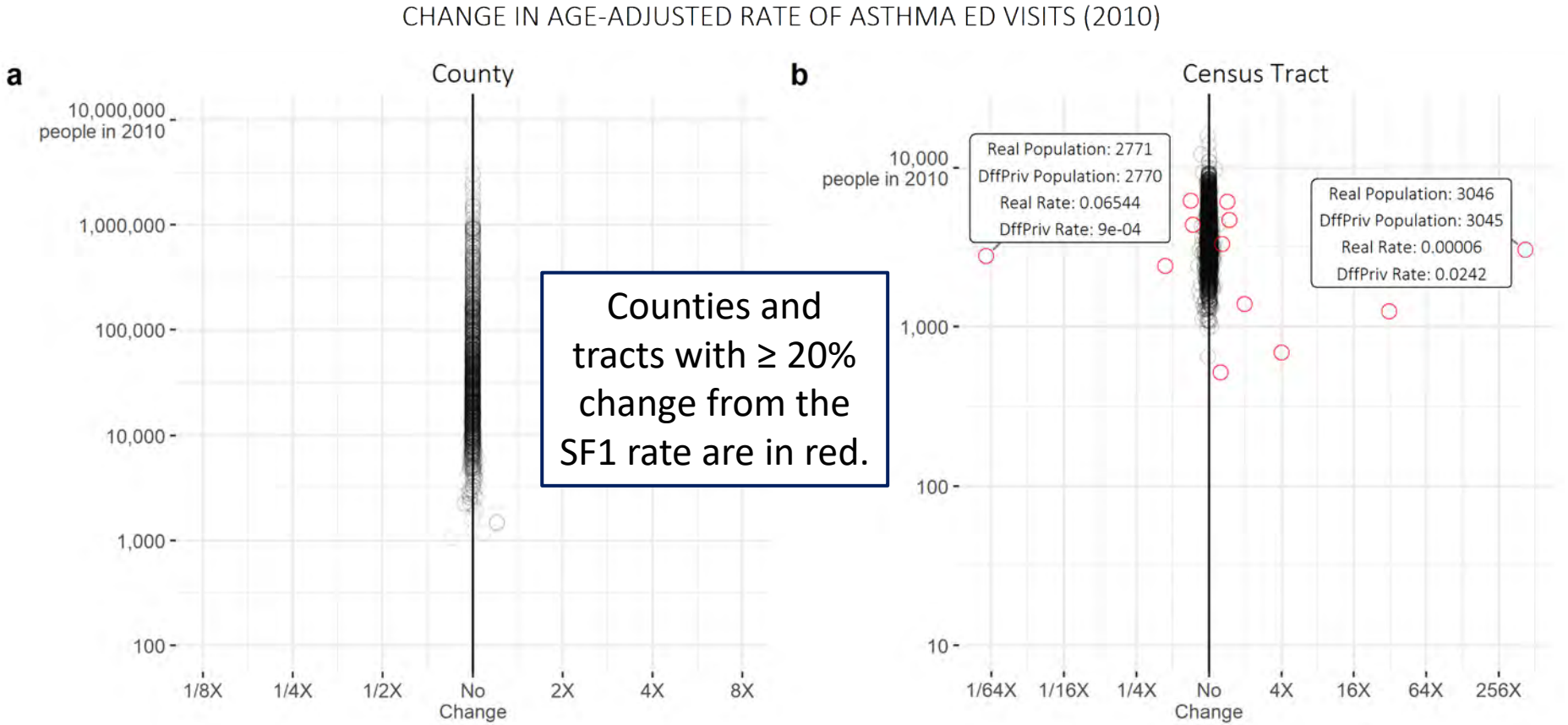
Methods

- Calculated age-adjusted rates with and without DP
- Value of percent difference calculated



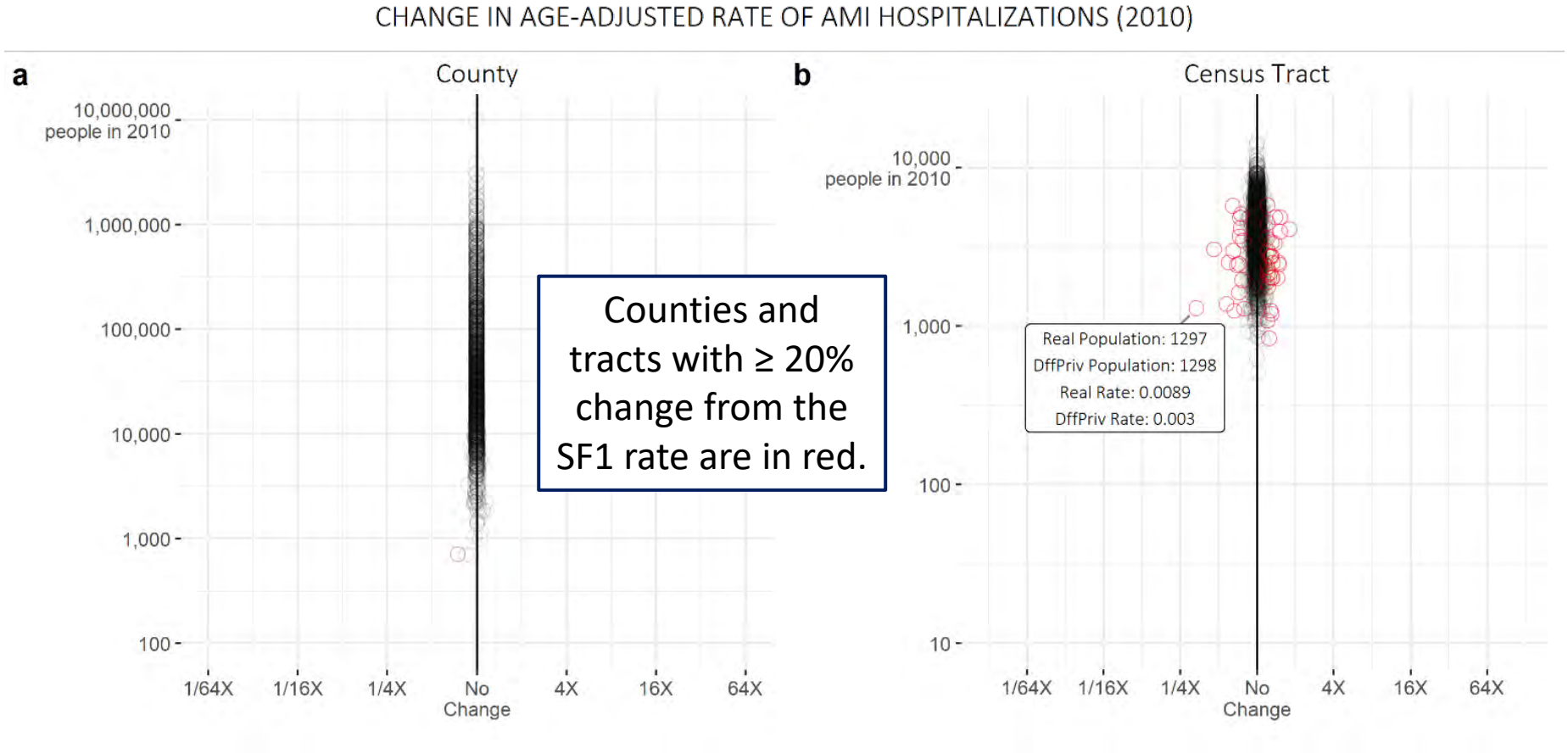
- Evaluated changes in rates between 2000 and 2010 using SF1 and DP denominators

Hosp/ED

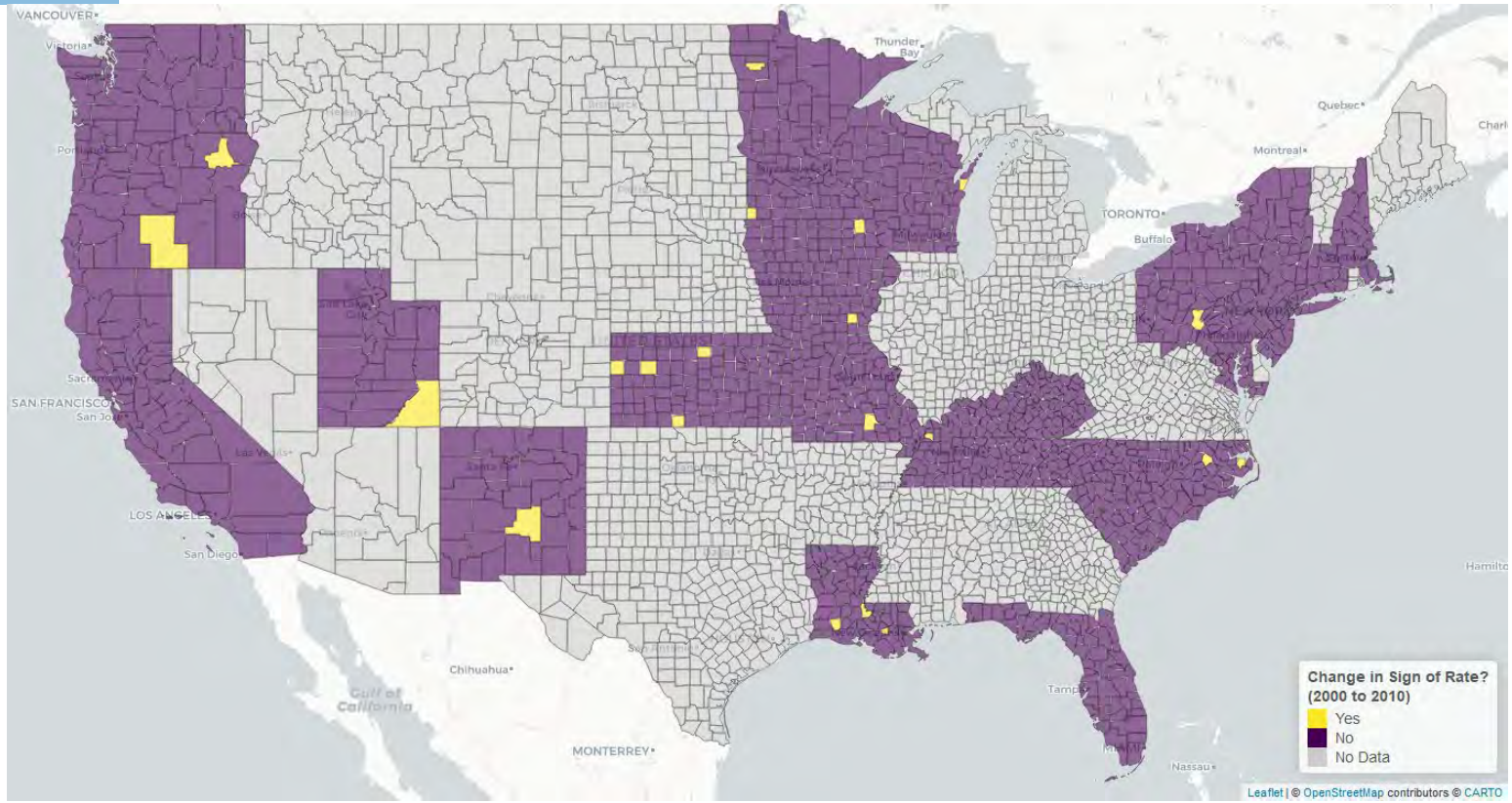


Counties and tracts with ≥ 20% change from the SF1 rate are in red.

Hosp/ED

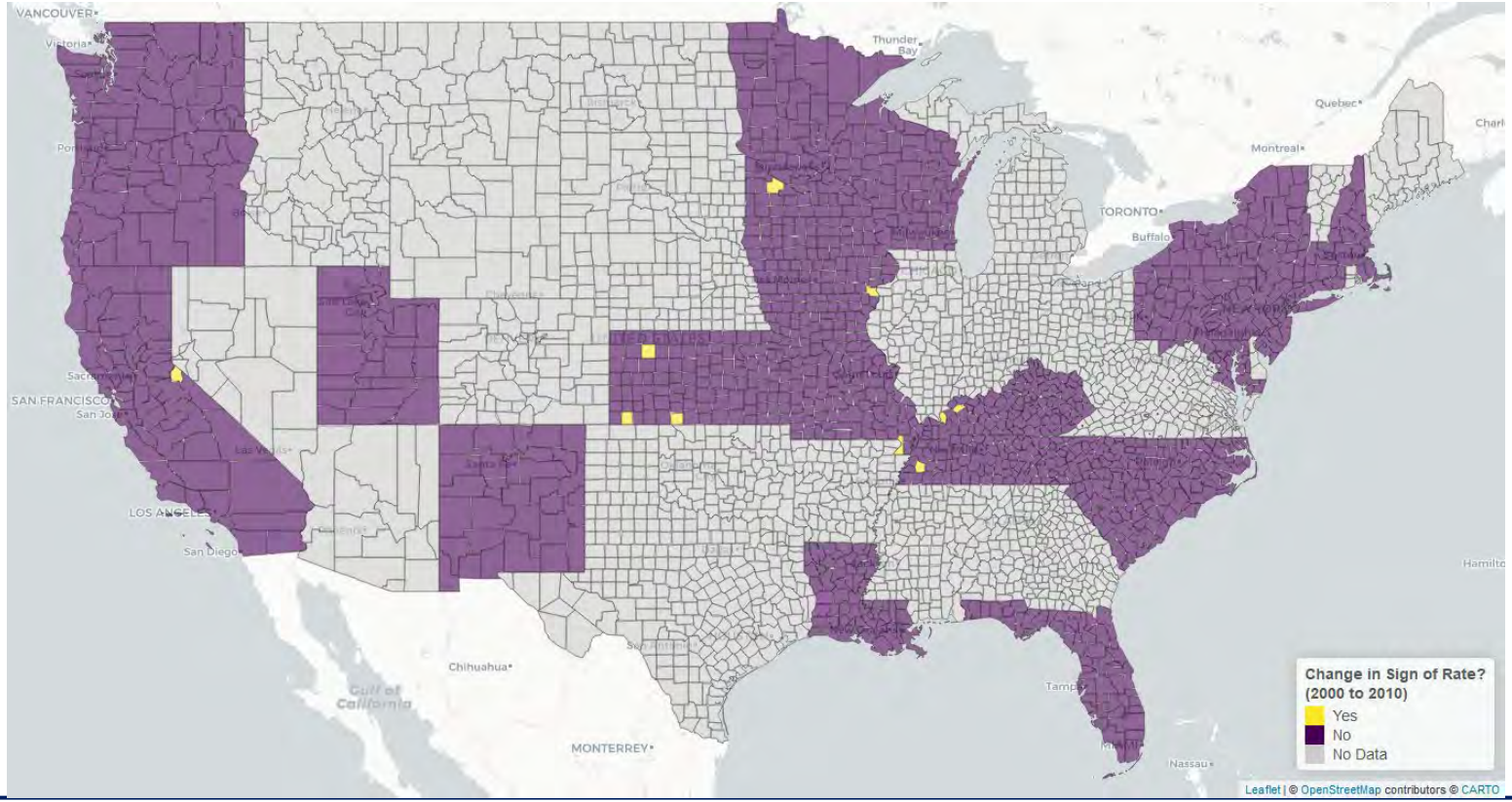

CHANGE IN AGE-ADJUSTED RATE OF AMI HOSPITALIZATIONS (2010)
a
County
10,000,000 people in 2010
1,000,000
100,000
10,000
1,000
100
1/64X
1/16X
1/4X
No Change
4X
16X
64X
b
Census Tract
10,000 people in 2010
1,000
100
10
Real Population: 1297
DffPriv Population: 1298
Real Rate: 0.0089
DffPriv Rate: 0.003
Counties and tracts with ≥ 20% change from the SF1 rate are in red.

Hosp/ED



Counties where the age-adjusted rates of asthma emergency department visits from 2000 to 2010 changed signs due to differential privacy are shown in yellow.

Hosp/ED



Counties where the age-adjusted rates of acute myocardial infarction hospitalizations from 2000 to 2010 changed signs due to differential privacy are shown in yellow.

Block: Assessing the impact on the Community Assessment for Public Health Emergency Response (CASPER)

Data and Methods

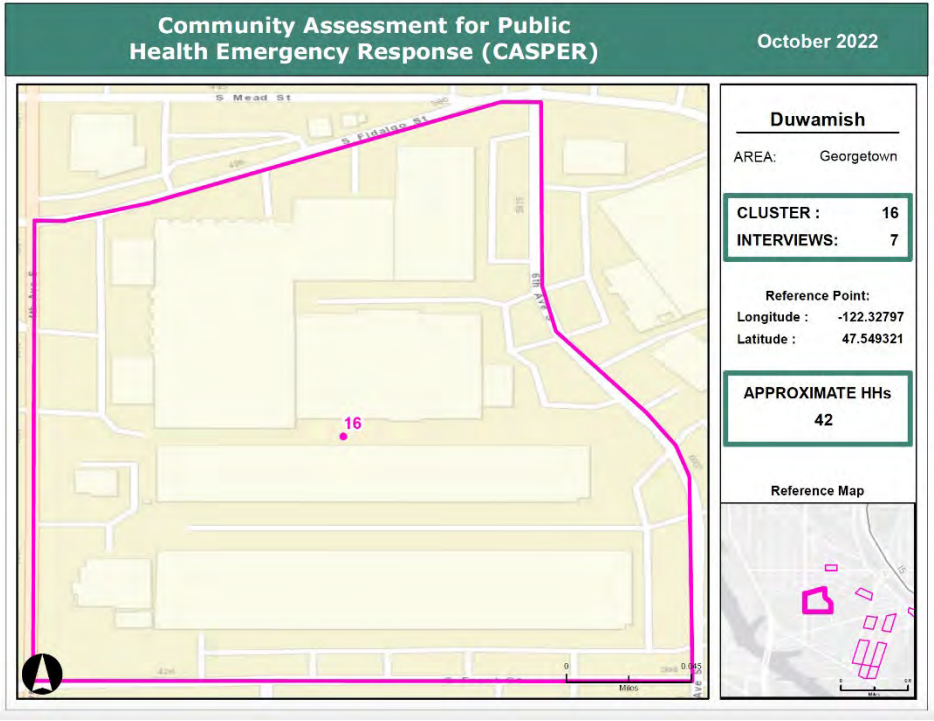
Data

- **2020 Census data at the block level**
  - Occupied households (H3)
- **Uses additional data, where needed**
  - Total population (P1)
  - Population in households by age (P16)
  - Race of householder (H6)
  - Hispanic or Latino origin of householder by race of householder (H7)

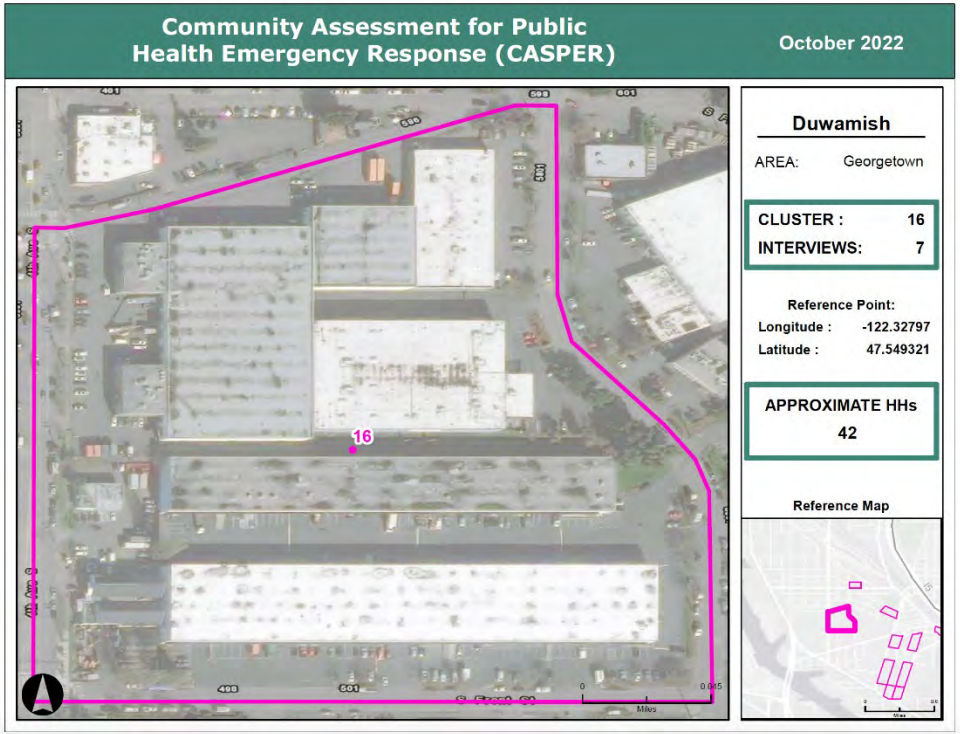
Methods

- Block-level data necessary for selecting sample and providing maps to send staff to the field
- Census data used to select 30 “clusters” (blocks) with probability proportional to size
- Seven households interviewed within each cluster
- Survey data are weighted based on 2020 Census data to gain accurate estimates of needs

CASPER



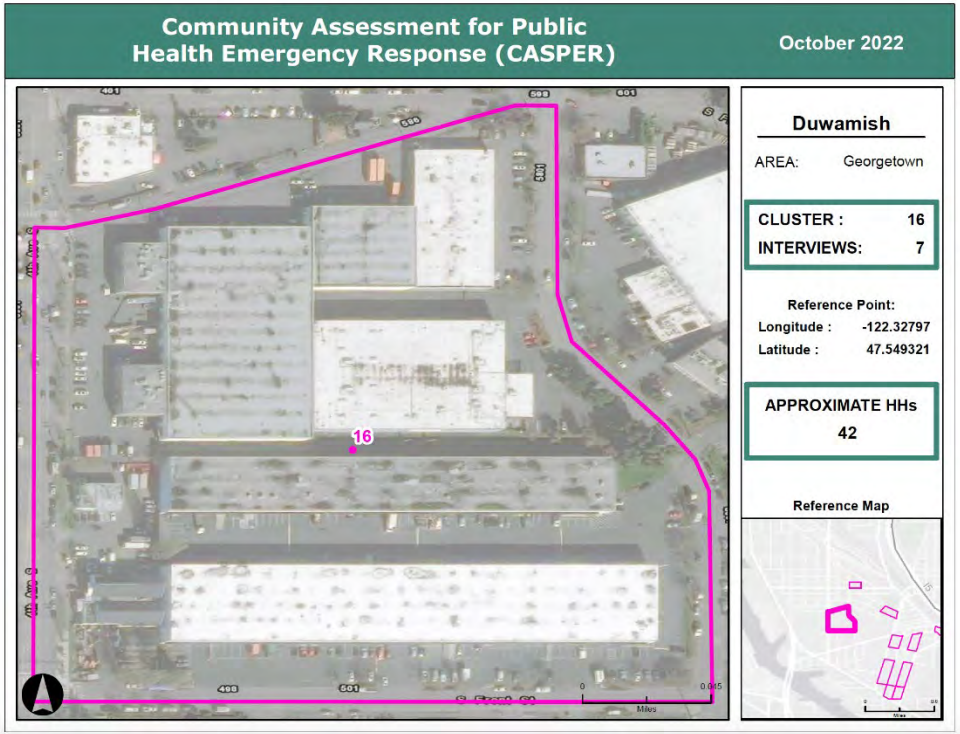
Sample cluster (#16) showing the approximate number of households (42) in the block according to 2020 Census data



Satellite imagery of the cluster (#16) where 7 interviews should have occurred through systematic selection of the 42 households

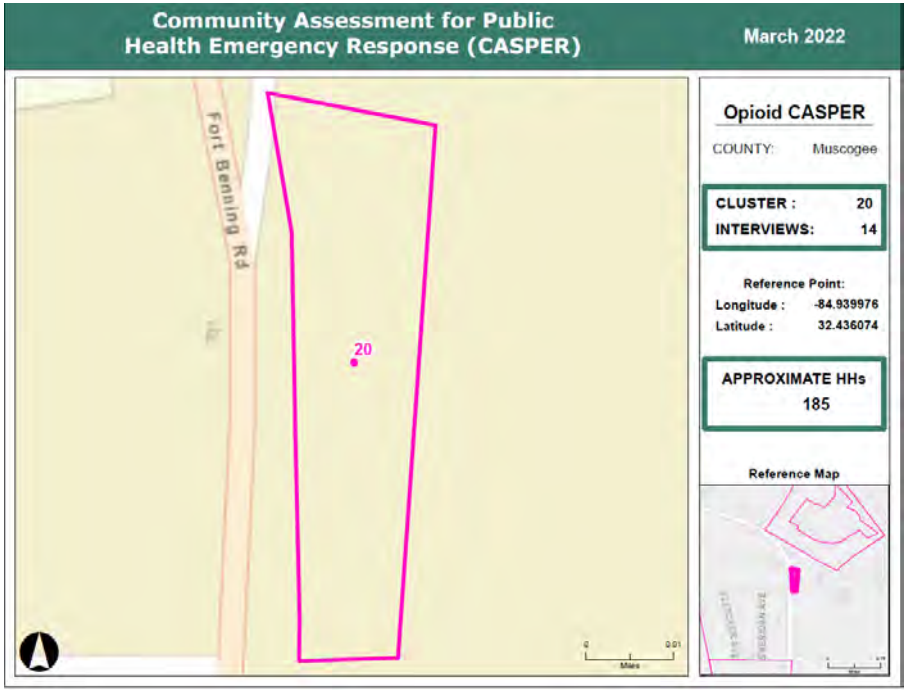
CASPER

- **No households actually in the census block**
  - Assumption that one of the buildings was an apartment and could have accounted for the 42 households
  - No households – only a food bank
- **Interviewed a few persons experiencing homelessness**

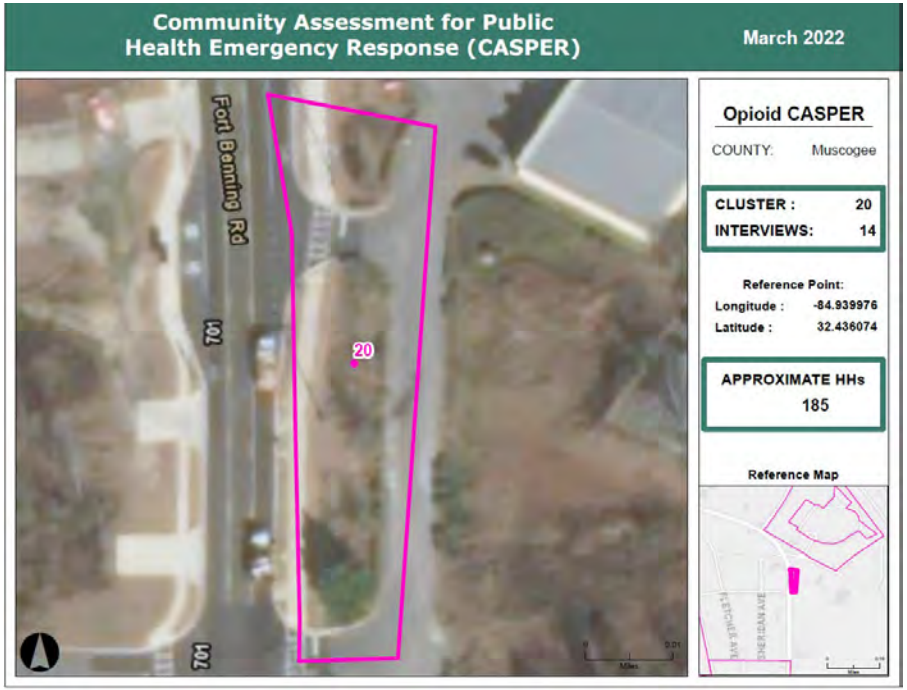


Satellite imagery of the cluster (#16) where 7 interviews should have occurred through systematic selection of the 42 households

CASPER



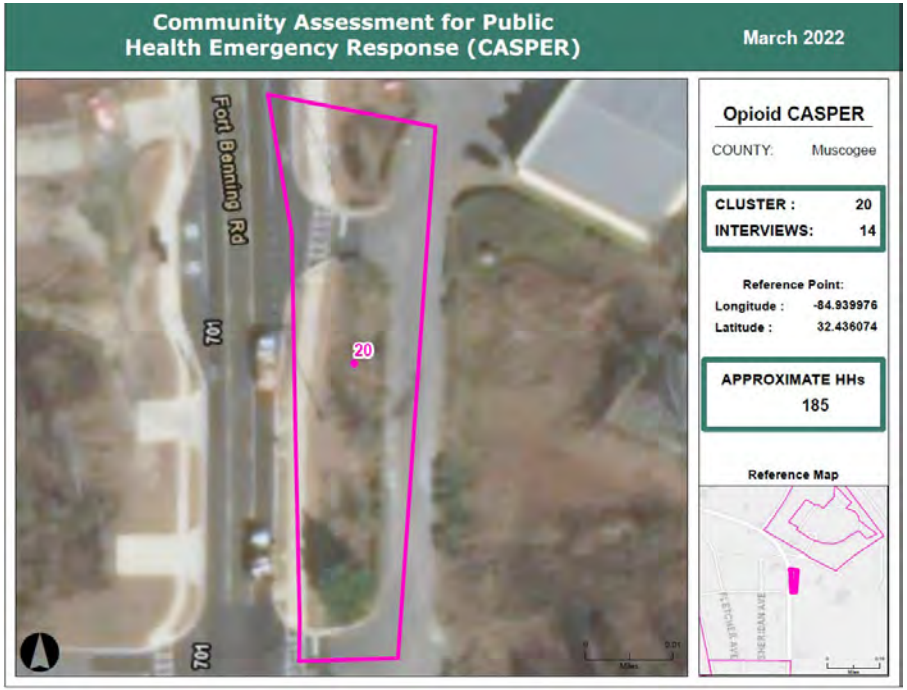
Sample cluster (#20) showing the approximate number of households (185) in the block according to 2020 Census data



Satellite imagery of the cluster (#20) where 14 interviews should have occurred through systematic selection of the 185 households

CASPER

- **Teams arrived at cluster expecting 185 households and no households were available to interview**
- **Invalidated the entire assessment**
  - At least one household must be interviewed in each cluster
  - A minimum of 168 households must be interviewed overall
- **CASPER could not be completed after several months of planning and effort**



Satellite imagery of the cluster (#20) where 14 interviews should have occurred through systematic selection of the 185 households

Conclusions From CDC Use Cases

County

- Data showed some overall improvement in age-adjusted rates*
- Significant differences in rates, particularly in counties with smaller populations and when stratifying age-adjusted rates
- Data remain problematic for age-specific rates

**no stratification by sex or by race/ethnicity or urban/rural or any sub-population*

Census tract

- Data remain problematic when calculating age-adjusted rates
- Changes in age groups impact overall rate calculations–even if total population counts don't change significantly

Block

- Data are problematic for characterizing risks accurately and identifying / targeting populations at higher risk
- Can not characterize communities accurately, particularly for situational awareness and emergency response activities

Public Health Implications

Under/overestimation of disease, disparity, or healthcare use rates

Trend monitoring over time

- Detecting true increases or decreases
- Setting targets
- Implementing evidence-based interventions
- Monitoring progress

Finer spatial resolution & sub-populations data

- Disseminating more local-level data
- Targeting populations accurately
- Allocating CDC resources

Measuring baselines and progress of CDC health equity goals

Acknowledgments

- CDC Differential Privacy Interest Group
- CDC Tracking Program Differential Privacy Content Workgroup
- CDC Office of Science
- Mel Smith and Ryan Davis, Geospatial Research, Analysis, and Services Program
- Jane Henley, Division of Cancer Prevention and Control
- Amy Schnall, Health Studies Section
- National Center for Health Statistics (NCHS)
- IPUMS

Follow Tracking on social media:

facebook.com/CDCEPHTracking

@CDC_EPHTracking

ENVIRONMENTAL PUBLIC HEALTH TRACKING

Contact Me:

Angela Werner, PhD, MPH
awerner@cdc.gov

For more information, contact NCEH
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

CDC