Evaluating the Potential Impacts of Differential Privacy on Public Health

A Series of Use Cases from the Centers for Disease Control and Prevention (CDC)

Angie Werner, PhD, MPH

June 22, 2022

2020 Census Data Products:
Workshop on the Demographic and Housing Characteristics Files

Public Health Surveillance

"The ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice"

https://www.cdc.gov/training/publichealth101/surveillance.html

Public Health Calculations

Crude Rates
overall average rate of disease

$$= \frac{\text{total number of cases}}{\text{total population}} \times 100{,}000$$

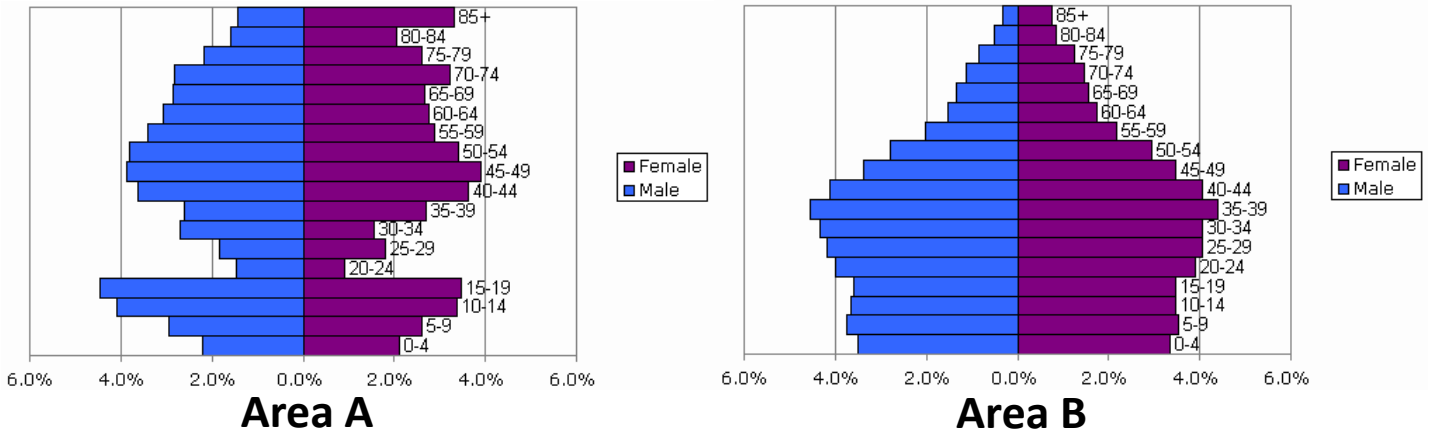
Age- or sex-specific rates
rate limited to a particular age group or sex

$$= \frac{\text{number of cases for 25–34-year-olds}}{\text{population of 25–34-year-olds}} \times 100{,}000$$

https://www.cdc.gov/csels/dsepd/ss1978/SS1978.pdf
https://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_StandardizedRates/EP713_StandardizedRates_print.html)

Age-standardized Rates

- Allow for comparing health outcomes among populations *with different age distributions*
- Applies standard age distribution to populations being compared
- Provides overall summary rate (“adjusted” rate)



https://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_StandardizedRates/EP713_StandardizedRates_print.html)
https://www.censusscope.org/us/s38/c37/chart_age.html and https://www.censusscope.org/us/s12/c95/chart_age.html

CDC Surveillance Systems*

Active Bacterial Core surveillance (ABCs)

FoodNet

Foodborne Diseases Active Surveillance Network (FoodNet)

BRFSS™

Behavioral Risk Factor Surveillance System (BRFSS)

WISQARS™

Web-based Injury Statistics Query and Reporting System (WISQARS)

USCS U.S. Cancer Statistics The Official Federal Cancer Statistics

U.S. Cancer Statistics

NSSP National Syndromic Surveillance Program BioSense Platform

National Syndromic Surveillance Program (NSSP)

ENVIRONMENTAL PUBLIC HEALTH TRACKING

National Environmental Public Health Tracking Program

NNDSS National Notifiable Diseases Surveillance System

National Notifiable Diseases Surveillance System (NNDSS)

NARMS National Antimicrobial Resistance Monitoring System

National Antimicrobial Resistance Monitoring System (NARMS)

PRAMS Pregnancy Risk Assessment Monitoring System

Pregnancy Risk Assessment Monitoring System (PRAMS)

USDSS United States Diabetes Surveillance System

National Diabetes Surveillance System (NDSS)

LINKING DATA TO SAVE LIVES

National Violent Death Reporting System (NVDRS)

VAERS

Vaccine Adverse Event Reporting System (VAERS)

Surveillance Resource Center A–Z Index | CDC *Partial list of CDC surveillance systems

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

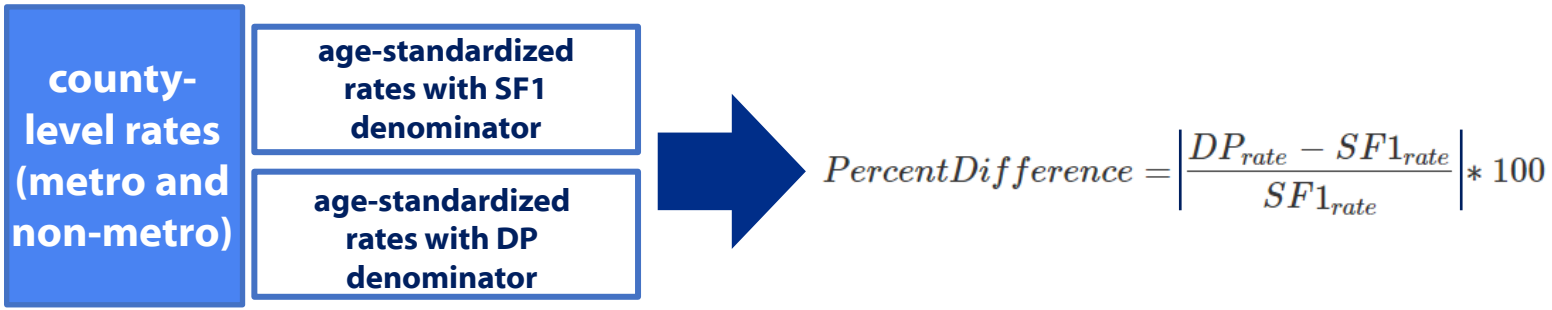
COVID-19

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 (v3-16-2022)

Methods



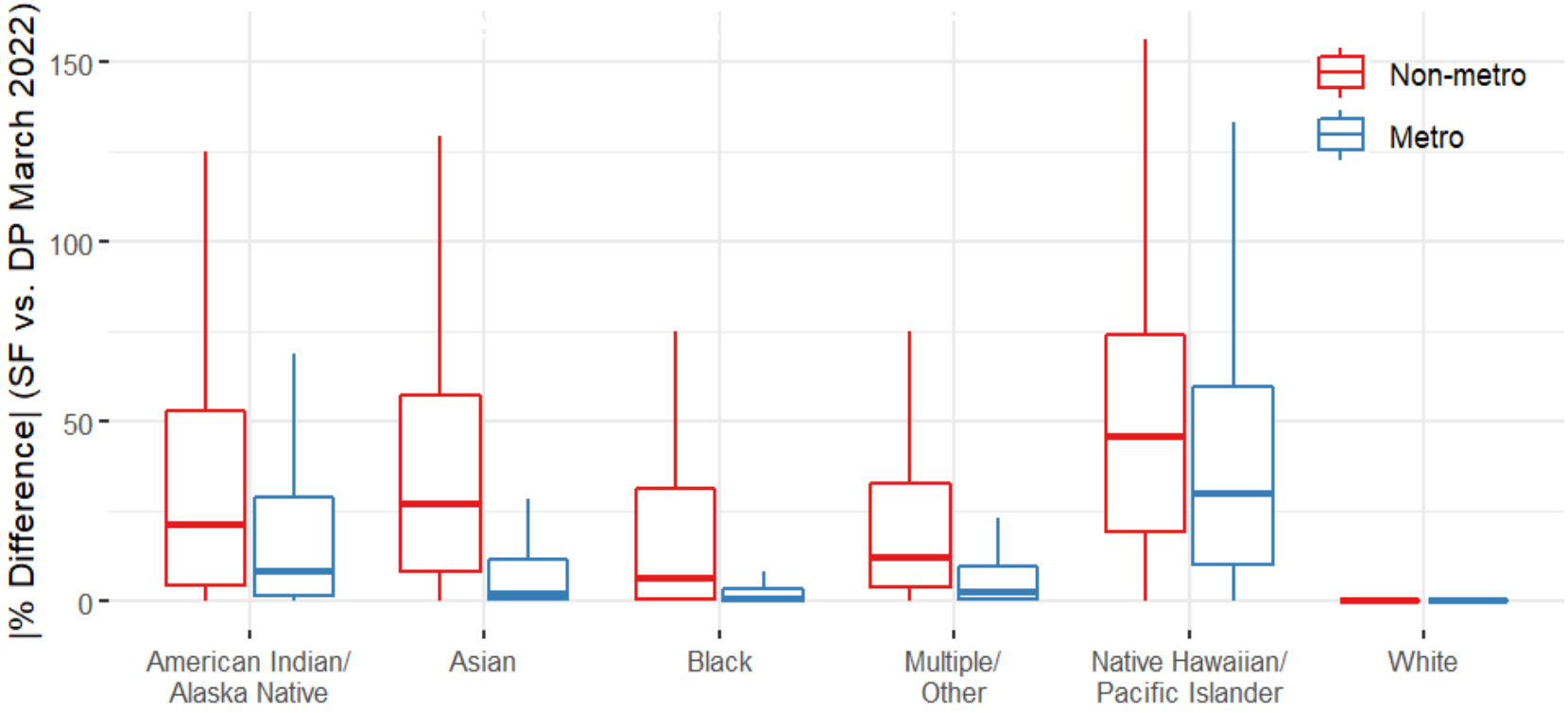
*included if case report date, age, race, and county of residence were submitted

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

COVID-19

Results

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



Source: US Census Bureau 2022; CDC Environmental Public Health Tracking Program

COVID-19

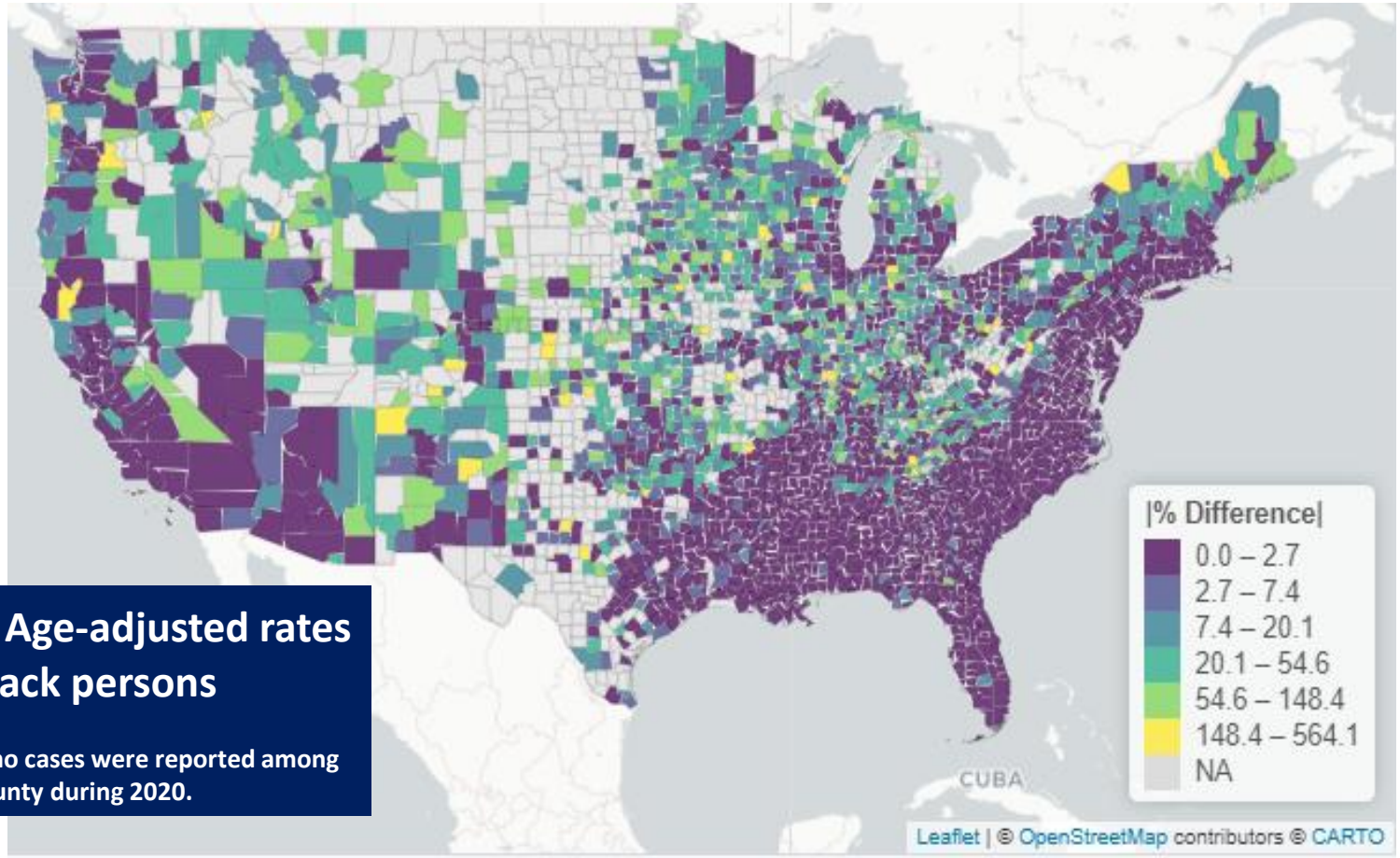
Results

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

Race/Metro or Non-Metro	Median %	Maximum %
AI/AN – Non-metro	21.4	900.0
AI/AN – Metro	8.3	642.5
Asian – Non-metro	27.1	900.0
Asian – Metro	2.3	631.1
Black – Non-metro	6.4	500.0
Black – Metro	0.4	564.1
Multiple/Other – Non-metro	12.4	700.0
Multiple/Other – Metro	2.8	317.6
Native Hawaiian/Other Pacific Islander – Non-metro	45.7	1000.0
Native Hawaiian/Other Pacific Islander – Metro	29.9	799.5
White – Non-metro	0.1	79.5
White - Metro	0.0	6.2

COVID-19

Results



|% 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.

COVID-19

Implications

- Use of DP in denominator data could distort CDC's health equity efforts
  - ➢ Unable to distinguish between real increases or decreases in incidence
- Greater impacts on smaller populations (e.g., racial minority groups, rural communities)
- Impacts on existing health equity challenges
- Could result in incorrectly targeting interventions or improperly allocating scarce medical resources

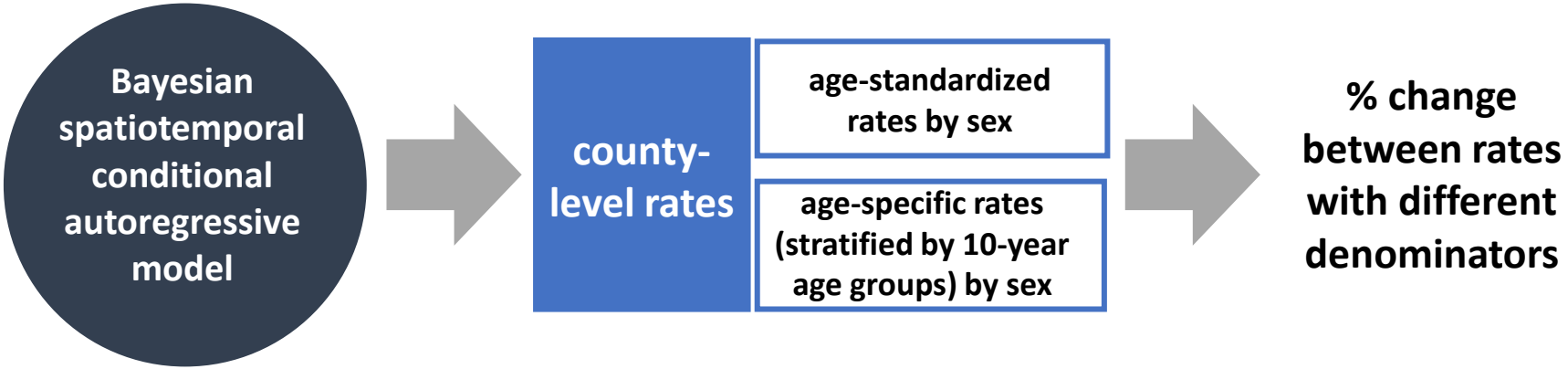
County level: Assessing the impact on heart disease mortality overall and by sub-group

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 (v3-16-2022)

Methods



Heart Disease Mortality

Results

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

*Outliers truncated.

Results

- Overall age-standardized rates
  - Estimates created with DP denominators almost always within 20% of those using the NCHS bridged-race denominator
  - Those areas with >20% difference generally occurred in areas with smaller populations (<10,000 people)
- Rates stratified by age group and sex
  - Almost half (43.3%) of rates had more than 20% difference
  - Almost all estimates with larger differences occurred in smaller populations (<10,000 people)
- DP tended to shift estimates so higher estimates more common

Heart Disease Mortality

Implications

- Models critical for cardiovascular disease public health surveillance
- Impacts particularly on small populations – including rural communities and racial/ethnic minority groups
- Estimates used to identify what counties have highest burden
- Used to target geographically focused interventions
- Rates generally inflated with DP denominator – counties could be less likely to meet certain goals (e.g., HealthyPeople2030) because of introduced error

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

Data and Methods

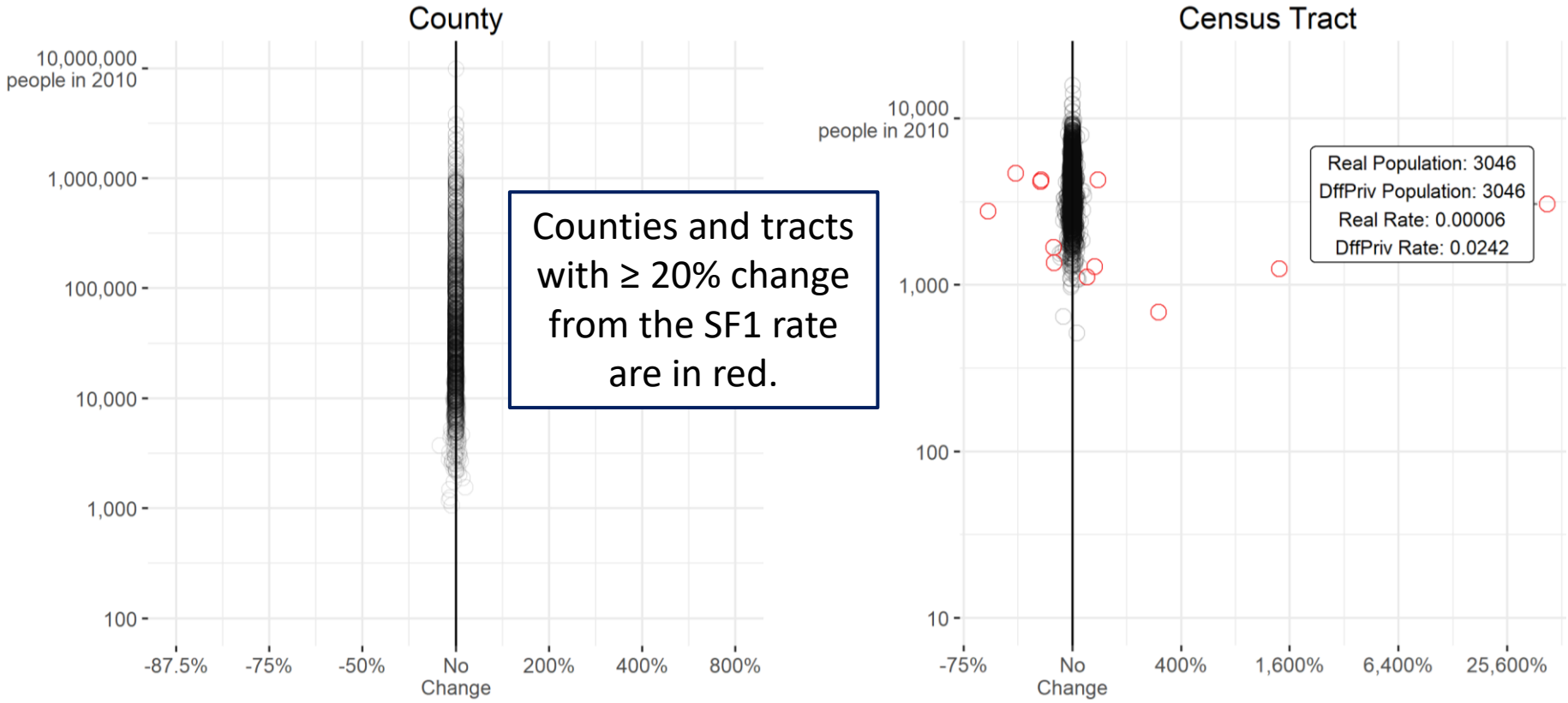
Data

- **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 (v3-16-2022)

Methods

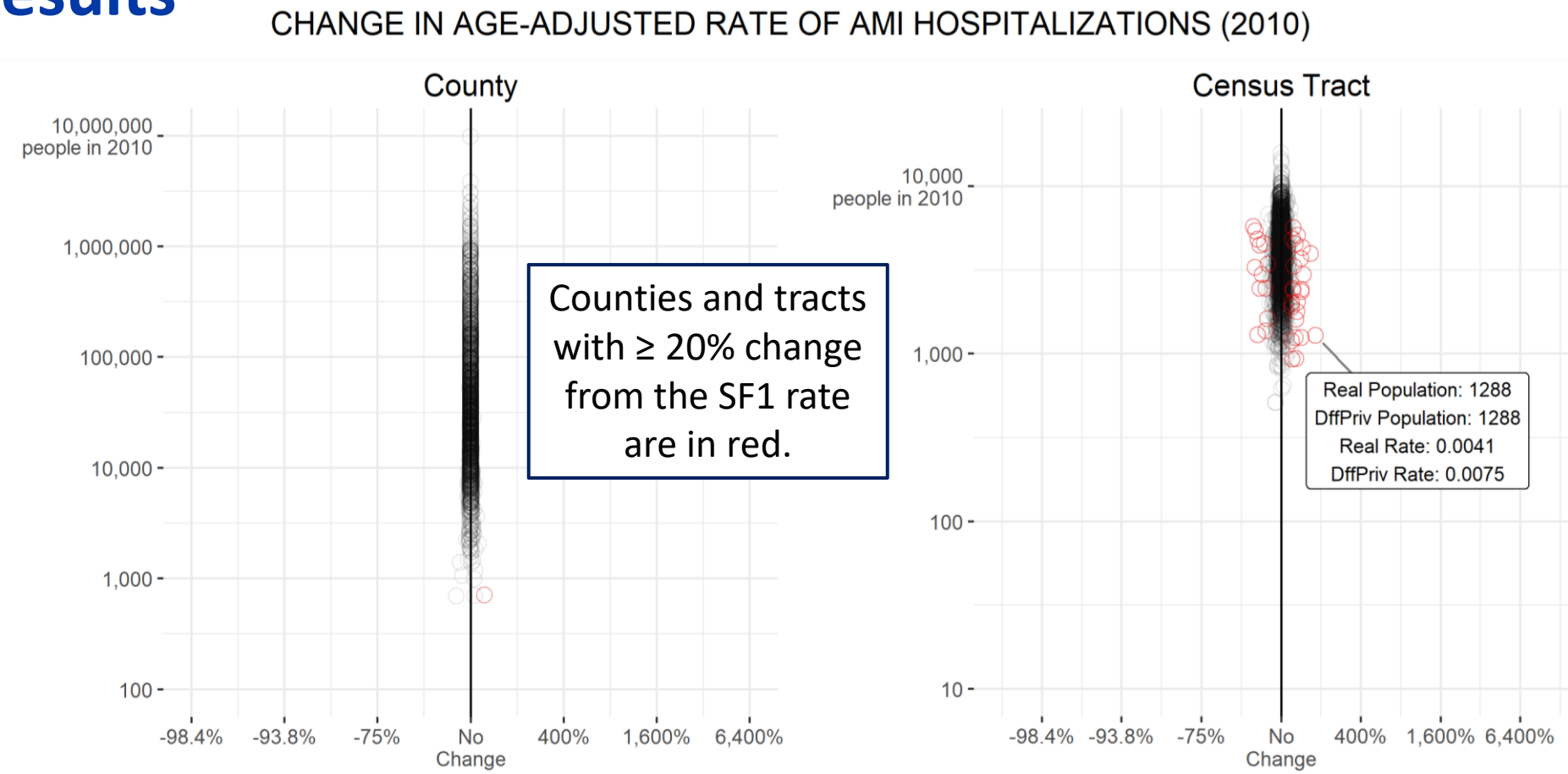
- Calculated age-adjusted rates with and without DP
- Value of percent difference calculated

Results


CHANGE IN AGE-ADJUSTED RATE OF ASTHMA ED VISITS (2010)
County
10,000,000 people in 2010
1,000,000
100,000
10,000
1,000
100
-87.5%
-75%
-50%
No Change
200%
400%
800%
Counties and tracts with ≥ 20% change from the SF1 rate are in red.
Census Tract
10,000 people in 2010
1,000
100
10
Real Population: 3046
DffPriv Population: 3046
Real Rate: 0.00006
DffPriv Rate: 0.0242
-75%
No Change
400%
1,600%
6,400%
25,600%

Hospitalizations and ED Visits

Results



Results

County level

- Rates improved when compared to the previous version (May 2020) of the DP demonstration data

Census tract level

- Changes exceeded 5% in certain tracts
- Large differences in rates
- Age-adjusted rates sensitive to population changes within individual age groups (even if total population count did not typically change)

Implications

- Surveillance systems working to obtain timelier, finer spatial resolution data
- Impedes ability to obtain more local level public health data and calculate accurate rates
- Age-adjusted rates sensitive to population changes within individual age groups (even if total population count did not typically change)

Block level: Assessing populations living near environmental hazards

Data and Methods

Data

- Block-level data by race, ethnicity, age, sex
- 2010 Decennial Census published data and 2010 DP demonstration data (v3-16-2022)

Methods

Generated demographics & maps → Enumerated population based on Census variables → Used area proportion technique → Estimated* population values in blocks not completely within designated buffer

Accuracy of area proportion technique directly related to geospatial granularity of data

*estimate calculated based on proportion of area of block included within the distance

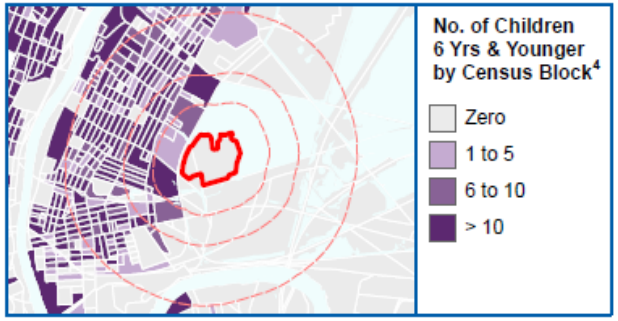
Environmental Hazards

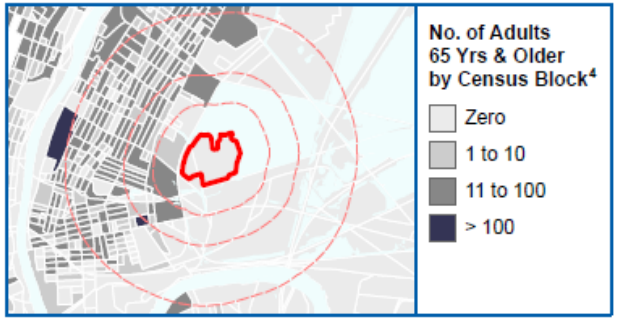
Results

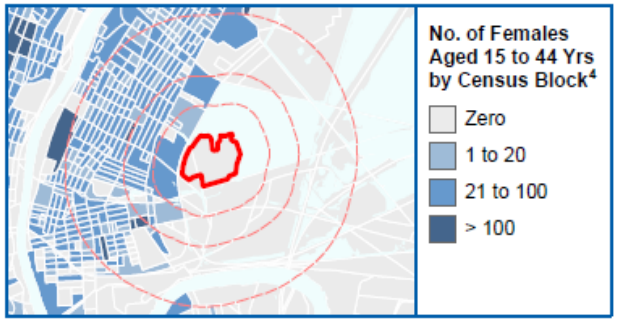
Keegan Landfill site population maps

Variation not significant enough to change classification assignments for blocks within one mile of site.

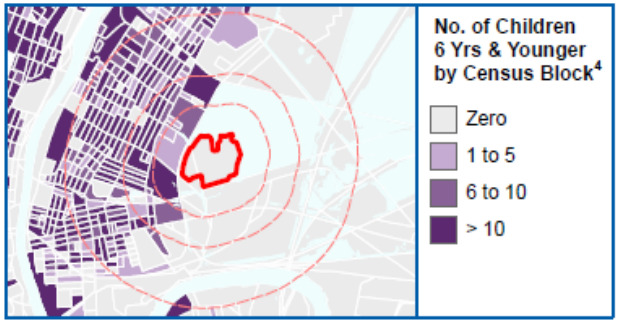
2010 Published Data

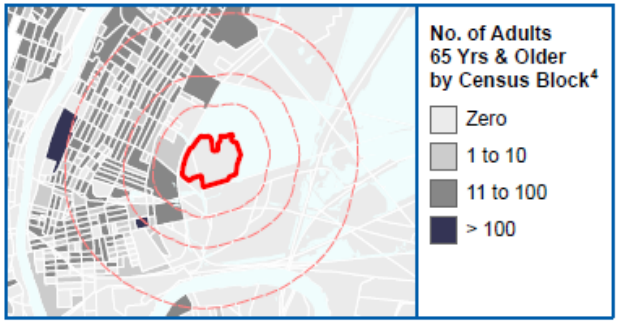


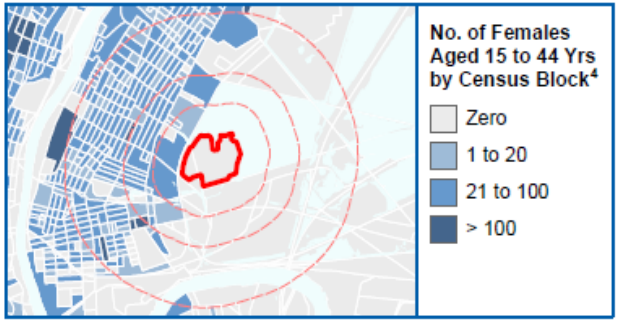




2010 Demonstration Data







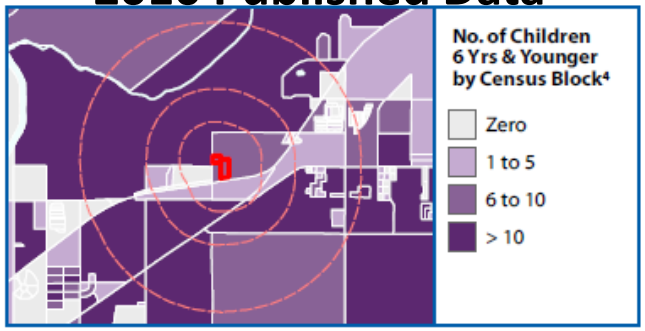
Environmental Hazards

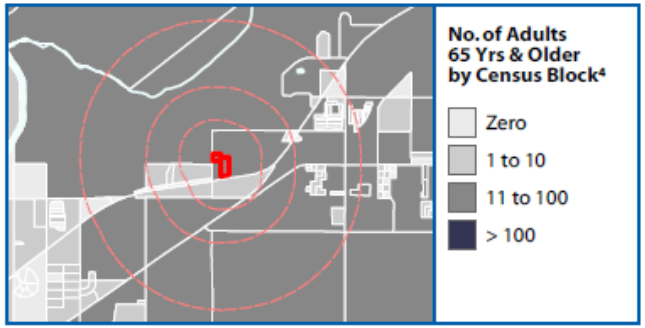
Results

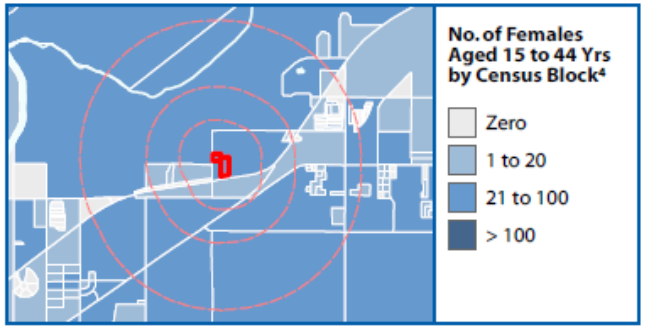
Arkla Terra Superfund site population maps

Variation <u>was</u> significant enough to change classification assignments for blocks within one mile of site.

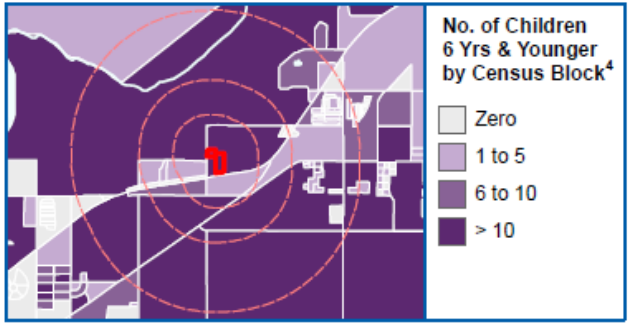
2010 Published Data

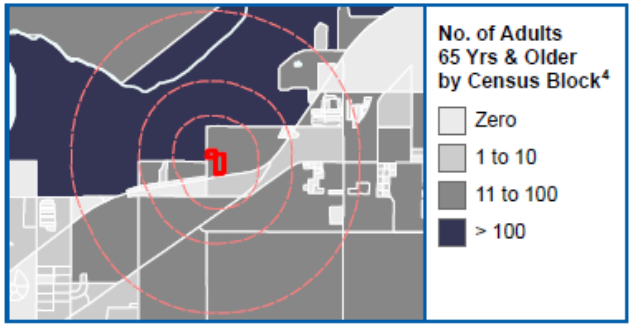


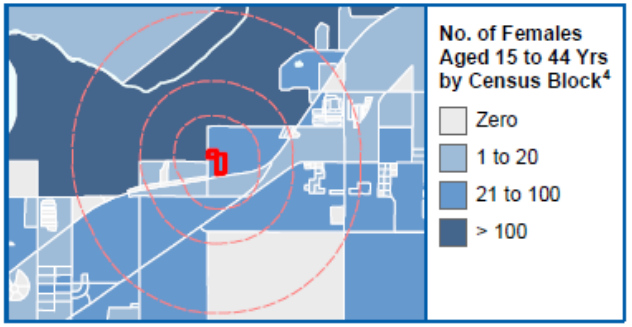




2010 Demonstration Data







Results

Range of 0.1%–36.4% difference in demographic estimates

Densely populated site

Measure	Difference
Black Alone	-7.6%
Am. Indian & Alaska Native Alone	-14.9%
Native Hawaiian & Other Pacific Islander Alone	+36.4%

Higher percentage change; range of 1.3%–100% difference

Low-density population site

Measure	Difference
Black Alone	+5.4%
Am. Indian & Alaska Native Alone	-30.4%
Native Hawaiian & Other Pacific Islander Alone	-100.0%
Some Other Race Alone	-15.1%
Two or More Races	-8.5%
Hispanic or Latino (of any race)	-9.7%

Implications

- Exposure assessments require accurate information
- Rely on this information to better identify and assess risk to general population and people who are at higher risk for conditions associated with exposure
- DP allows for potential for mischaracterization of risks due to inaccurate placement of populations who are at higher risk
- Underestimation of population due to DP may inadvertently exempt affected areas from consideration for assessment

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 will be problematic for characterizing risks accurately and identifying / targeting populations at higher risk

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

- **Adam Vaughan**, Division for Heart Disease and Stroke Prevention
- **Mel Smith** and **Ryan Davis**, Geospatial Research, Analysis, and Services Program
- **Nick Skaff**, National Environmental Public Health Tracking Program
- CDC Differential Privacy Interest Group
- CDC Tracking Program Differential Privacy Content Workgroup
- CDC Office of Science
- National Center for Health Statistics (NCHS)
- IPUMS

Angie Werner, PhD, MPH
awerner@cdc.gov

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