UNCERTAINTY IN DEMOGRAPHIC AND SOCIOECONOMIC DATA THE USE OF DIFFERENTIAL PRIVACY FOR DISCLOSURE CONTROL, AND ITS POTENTIAL IMPACT ON AGE AND **RACE/ETHNICITY COUNT**

DAVID VAN RIPER, UNIVERSITY OF MINNESOTA SETH SPIELMAN, MICROSOFT

SOURCES OF RACE/ETHNICITY DATA

American Community Survey

- ACS is a sample of housing units.
- 3.5M housing units are sampled, about 2M responses are collected (~60% response rate).
- ACS population counts are estimates, from a sample, and and as a result carry uncertainty.

Decennial Census

- The Decennial Census is complete enumeration of the US population.
- Published data have always obfuscated responses to prevent reidentification.
- For 2020 Census is adopting a formal privacy framework which injects noise into the data.

ACS ESTIMATE QUALITY: HISPANIC POP BY TRACT



Margin of error 100% of the estimate

15% of all Census Tracts have a margin of error that is 100% of more of the estimate.

Margin of error 50% of the estimate

53% of all Census Tracts have a margin error that is 50% of more of the estimate.

Margin of error 10% of the estimate

99% of all Census Tracts have a margin of error that is 10% of more of the estimate.

Note: Zero estimates included in figure but excluded from % of tracts > than 100%, 50%, 10% thresholds Figures truncate outliers, excluding the largest 1% of areas

ACS ESTIMATE QUALITY: TOTAL POP BY TRACT



Margin of error 100% of the estimate

Less than 1% of Census Tracts have a margin of error that is 100% or more of the estimate.

Margin of error 50% of the estimate

Less than 1% of all US Census Tracts have a margin error that is 50% of more of the estimate.

Margin of error 10% of the estimate

35% of all US Census Tracts have a margin of error that is 10% of more of the estimate.



African-American Population

Margin of Error and Estimate

Margin of Error and Estimate Native Hawaiian/Pacific Islander: All US Census Tracts 2019 ACS

700

ACS ESTIMATE QUALITY: HISPANIC POP BY COUNTY



Margin of error 100% of the estimate

3% of all counties have a margin of error that is 100% of more of the estimate.

Margin of error 50% of the estimate

10% of counties have a margin error that is 50% of more of the estimate.

Margin of error 10% of the estimate

13% of all counties have a margin of error that is 10% of more of the estimate.



SOLUTIONS

· Goldilocks geographies.

- Census tracts can be too small to provide reliable racial/ethnic group estimates. Counties can be too large, meaningful intra metropolitan socio-spatial variation is lost.
- Website: <u>https://reducinguncertainty.org/</u>
- We've developed <u>software</u> to generate "optimal" geographies based on user defined constraints (margin of error, population size).
- Geodemographic Data
 - <u>Paper</u>, <u>Data</u>, <u>code</u>, <u>Interactive Map</u>
 - Contextual as opposed to a "variable" based approach.

SOLUTIONS



DECENNIAL 2020: STILL A MOVING TARGET

STATUS OF 2020 DECENNIAL DATA

- PL94-171 redistricting data
 - Released in August 2021
- Demographic and Housing Characteristics (formerly SF1)
 - Tentative plan for release in mid-to-late 2022
 - Delay of one year compared to prior decennials
- Detailed Demographic and Housing Characteristics (formerly SF2)
 - No current timeline for release
 - Typically released six months after SF1
- Why the delays?

NEW DEMONSTRATION DATA

- Census Bureau will release another demonstration product sometime soon
- Will contains tables proposed for the DHC (formerly SF1)
 - Sex by age by race / ethnicity
 - Household type / structure
 - Household relationship
 - Housing tenure

- New census block tables to the DHC
 - Sex by 5-year age bins by race / ethnicity for many more race / ethnicity combos
- Reduction in geographic detail
- Elimination of tables

Reduction in geographic detail

- Shift from census block or tract to county
- County variation > block or tract variation
 - Loving, TX (64) to Los Angeles, CA (10 million)
 - 156,250 times difference in population!
- 69 tables move from block or tract to county

Elimination of tables

- Current proposal eliminates 169 tables from SF1 / DHC
- 67 tables out of 71 in SF2 / DDHC may be eliminated

Elimination of tables

- Single year of age by sex by race by ethnicity
- Hispanic origin by specific origin
- Allocation counts

CONCLUSION

- From ACS and Decennial one can expect good city-level rates/population estimates
 - But within-city or county is harder to understand
- Geographic and demographic resolution matter
 - Units with larger counts will be more accurate
 - Demographic groups with larger counts will be more accurate
- It is possible to process publicly released data to improve estimates.
 - Particularly for ACS data
 - Less certain about decennial data since Bureau will not publish error bounds

QUESTIONS OR FEEDBACK:

vanriper@umn.edu

seth.spielman@colorado.edu

Code and data:

https://github.com/geoss/cdc-ncvhs-covid-2021