

ICPSR

Census 2020: Privacy Protection and the Future of Public Use Data Products

Association of Population Centers
Fall meeting, October 2019

Maggie Levenstein
ICPSR Director



Dissemination plans for 2020

Reduced number of tabular files

All of which will use noise infusion for confidentiality protection

Census currently debating ϵ

Noise infusion is not the problem

More information and probably less distortion than prior approaches

Approach will reduce (rather than aggregate) noise as geography increases

Critical to ensure that the tables that are produced are those that are needed

Use cases include research, commerce, government planning, and policy making

What will be produced?

Apportionment Product

State

Redistricting File (P.L. 94-171)

Census block (April 1, 2021)

Demographic Profile

Place/minor civil division

Demographic and Housing Characteristics File (DHC)

Various

Congressional District Demographic and Housing
Characteristics File

New congressional districts

Demographic & Housing Characteristics

SF1 replacement

Geography: “Data will be available for all geographies at and above the lowest level they are released. In many cases this will be at the Census Block level.”

Content: “DHC will not include tables that provide counts by detailed race, Hispanic origin, tribes, and population by household/family types”

<https://www2.census.gov/cac/sac/meetings/2019-09/status-update-2020-census-data-products-plan.pdf>

What about a PUMS?

Any PUMS will be synthetic

You cannot, by definition, make a PUMS that is differentially private

Existing PUMS has confidentiality protections that we know distort the data. Question is how these new approaches compare to the old ones.

We have been promised code that can be studied to determine impact.

We have been promised unperturbed data in the FSRDC

Launching virtual pilot for Title 13 only data products

➤ Other statistical agencies are participating - not NCHS

What else is promised?

SF2, Detailed race/Hispanic with household data

- They clearly do not yet have a technology that will make these safe and high quality

What about ACS?

Census has said that it will not implement “formal privacy” until 2025

They have determined re-identification risk is lower than they feared

They do not know how to make a synthetic ACS

2020 Dissemination Census Advisory Committee

<https://www2.census.gov/cac/sac/meetings/2019-09/status-update-2020-census-data-products-plan.pdf> (Devine and Hollingsworth)

<https://www2.census.gov/cac/sac/meetings/2019-09/update-disclosure-avoidance-administrative-data.pdf> (Abowd and Velkoff)

More fun: <https://www.census.gov/library/video/2019/protecting-privacy.html>

Accessible but more technical:

<https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a>

ICPSR

**Click to add
section title**

