Data User Perspectives on Potential Changes to Data Products to Protect Privacy

Michael Davern, PhD
Executive Vice President
NORC at the University of Chicago

June 7, 2019

NORC
at the UNIVERSITY of CHICAGO

Users Have Invested Heavily in Current Census Data

Governments, businesses, nonprofits, and academics have invested billions of dollars and volumes of research in systems and tools that use Census data for:

- Reading, ingesting, and transforming data
- Creating software programs, statistical programs, APIs, and applications
- Creating microsimulation models, evaluating programs, monitoring need for a program, and calculating funding formulas for programs
- Understanding data quality issues and error structure of the current data that informs best use
- Creating curated data products for dissemination and resale

These uses create a complex web of workflows, tools, and documentation that rely on the current data

Changing Census Data Will Incur Significant Costs

Changes in Census data will force users to:

- Modify existing workflow, tools, and documentation
- Analyze how the infusion of errors will impact the specific use of data, for example:
  - Microsimulation models created by CBO, RAND, Urban, MPR, etc.
  - The analysis of new data will require new tools (e.g. SAS/R/Stata, or software for handling new data's error structure)
- Invest in quality assurance as errors will result during updating of workflow, tools, and documentation to adjust to new data

Changes in Census Data Will Reduce Their Quality

New Census privacy protections will result in lower-quality data at a time when surveys face:

- Competition from proprietary, poorly researched, and opaque data alternatives to transparent, well-understood data from the Census and American Community Survey (ACS)
- Increased costs and decreased response rates
- Growing public skepticism as data are seen as a political tool (e.g., the 2016 election, redistricting, citizenship)
- Decreasing Census/ACS data quality, which could help favor other sources of data

Is Change Needed? Example of ACS Microdata

Current ACS microdata have existing privacy protections that make direct matching to other data sources very difficult

- Survey variables responses differ significantly from other sources of data in record match studies
  - Market research and administrative data matches:
    - measurement errors and conceptual differences lead to differences
- Survey data differ significantly in re-interview studies
  - ACS housing items: Index of Inconsistency (IOI) ranges from 4 to 69% with a median of 26%
  - ACS person items: IOI ranges from 0 to 73.2 % with a median of 22.1%

Is Change Needed? Example of ACS Microdata

- Census allocates, edits, and imputes most ACS items
  - Upwards of 25% of survey responses allocated
    - Census could remove allocation flags from PUMS for further protection
  - Complex edits are used for key items such as poverty rate and labor force participation
- ACS uses traditional statistical disclosure limitation, including coarsening, suppressing, sampling, and swapping

Census should clearly demonstrate these current privacy protections are *not adequate* for the ACS microdata before making a major change

Results from 2010 putative re-identification test should be overgeneralized

Call for Transparency and Dialogue

What do we need to move forward?

- Develop a public timeline for changes to Census and ACS summary files and microdata
- Create public versions of new “noise-infused” summary files and microdata for Census and ACS
  - To date we have seen math and theoretical explanations
- Engage with the research community after we have had time to work with the data ***before*** making data changes permanent

Davern-Michael@norc.org

Thank You!

NORC at the UNIVERSITY of CHICAGO

insight for informed decisions™