

Confidentiality for American Community Survey Respondents

Rolando A. Rodríguez, U.S. Census Bureau

Presented to the 2023 ACS Data Users Conference

May 16, 2023

*Any opinions and conclusions expressed herein are those of
the author(s) and do not reflect the views of the U.S. Census Bureau.*

Overview

1. Status of formal privacy for sample surveys
2. Balancing data accuracy and disclosure avoidance (DA)
3. Overview of current ACS Public Use Microdata Sample (PUMS) DA
4. Envisioning new tiers of ACS public data access
5. Research into synthesis and validation
6. Commitment to transparency

There are no plans to apply the 2020 Disclosure Avoidance System (DAS) to the ACS

- The Census Bureau has had several cooperative agreements to understand how statistical agencies can adapt to formal privacy (FP)
- A current agreement¹ specifically focuses on FP for sample surveys
- Numerous issues related to statistical practices in sample surveys make application of FP frameworks difficult
- Changes to ACS based on FP can only happen once the scientific understanding has improved

The Census Balancing Act

- Title 13 of U.S. Code requires that Census Bureau prevent “any publication whereby the data furnished by any particular establishment or individual under this title can be identified”
- But the statistics we produce must be useful for stakeholders to make decisions
- Data releases cannot be perfectly confidential and perfectly accurate at the same time



Confidentiality concerns are here, now



- Modern computational power changes how we need to think about the risk of revealing respondent identities or data
- An ethical attack on the 2010 Census reidentified unique individuals in the smallest census blocks with high confidence
- Agencies like the Census Bureau need to reassess their current disclosure avoidance (DA) practices

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping
- Partial synthesis

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping
- Partial synthesis
- Perturbation

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping
- Partial synthesis
- Perturbation
- Top and bottom coding

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping
- Partial synthesis
- Perturbation
- Top and bottom coding
- Coarsening

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping
- Partial synthesis
- Perturbation
- Top and bottom coding
- Coarsening
- Subsampling

How do we protect ACS microdata now?

- Traditional DA frameworks take an ad hoc approach to risk
- The ACS has a mix of DA methods
- Among ACS products, the ACS Public Use Microdata Sample (PUMS) has the most varied techniques
- Swapping
- Partial synthesis
- Perturbation
- Top and bottom coding
- Coarsening
- Subsampling
- What are the effects on accuracy and disclosure risk?

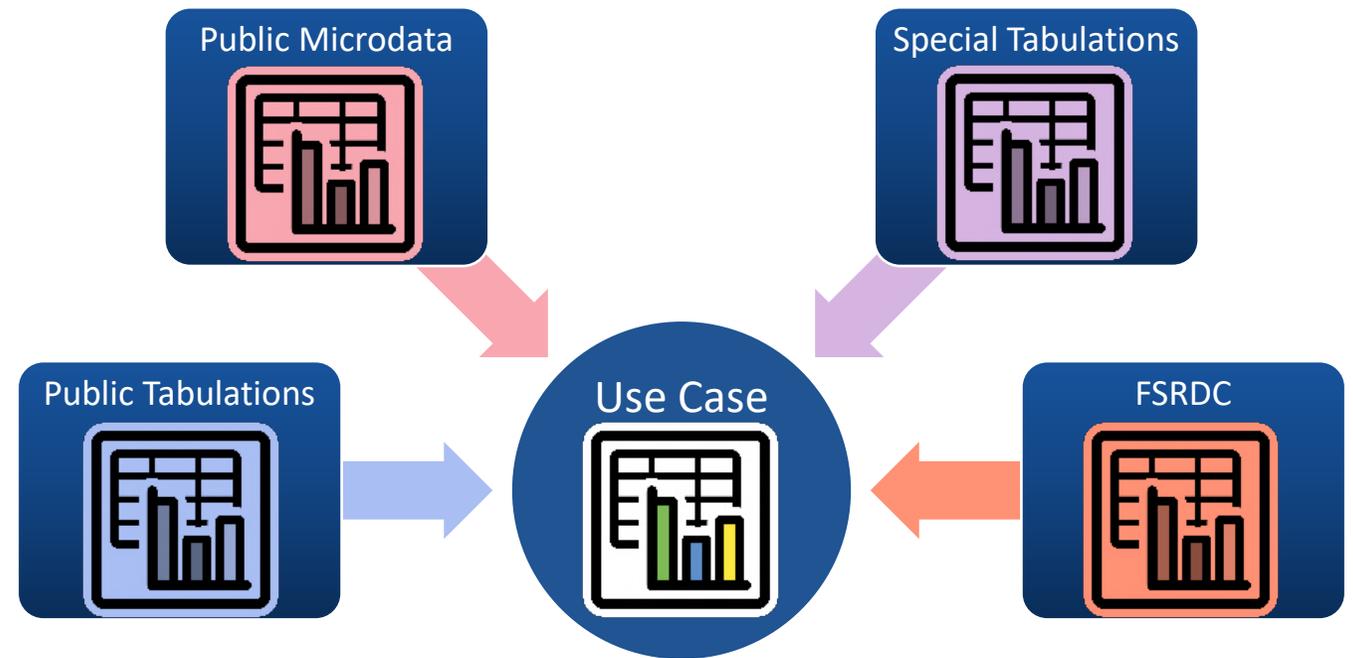
Ignoring the creeping uncertainty



- Each of the ACS PUMS DA interventions affects estimates and their errors
- Among the interventions, only subsampling has any specific weighting adjustments for the end user
- This means that PUMS analyses inherently undershoot total survey error
- Even with all these methods, there is no guarantee of confidentiality
- There must be a better way...

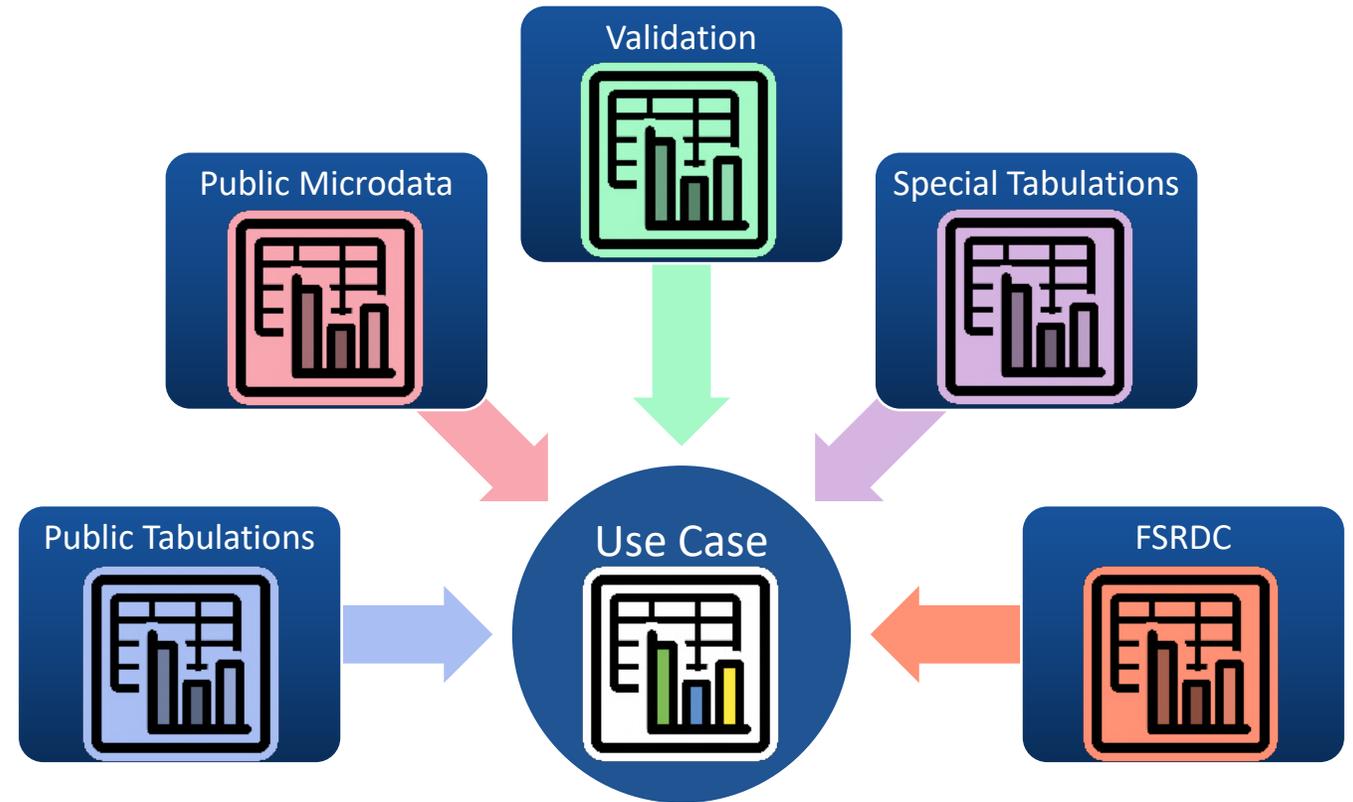
New tiers of access can benefit ACS users

- Tiered access allows for adaptive data publications that can support varying levels of accuracy and confidentiality
- Tiers like microdata allow for broad access and exploratory analysis but have increased protections and larger errors
- Tiers like the FSRDCs restrict access but allow for fine-tuned disclosure avoidance and error adjustments

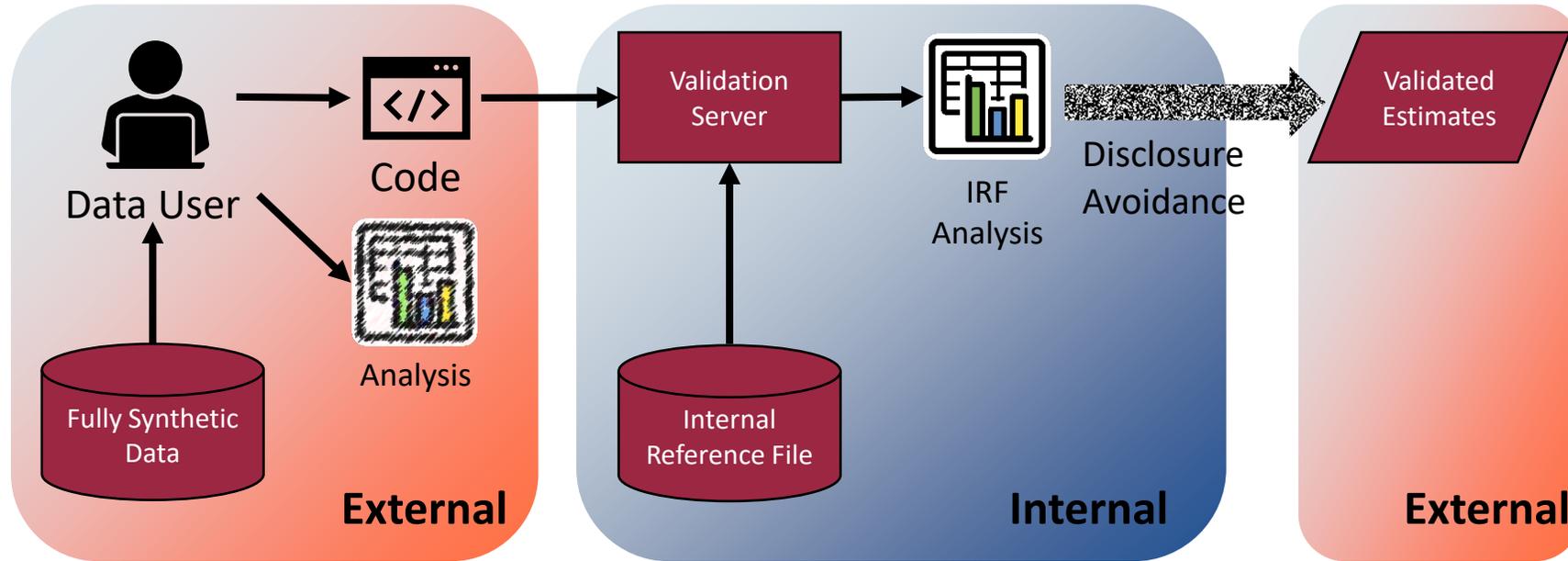


New tiers of access can benefit ACS users

- Tiered access allows for adaptive data publications that can support varying levels of accuracy and confidentiality
- Tiers like microdata allow for broad access and exploratory analysis but have increased protections and larger errors
- Tiers like the FSRDCs restrict access but allow for fine-tuned disclosure avoidance and error adjustments
- Adding validation would provide many of the benefits of FSRDCs to the entire ACS user community



Synthesize and validate



- Fully synthetic data protects respondents by sending all data through statistical models, meeting our mandate to produce only **statistical** results
- Validation allows analysis on pristine internal data, with outputs protected by DA methods specific to the type of output

Advantages over current PUMS

- 👍 Validated results run on internal data unaffected by disclosure avoidance
- 👍 Disclosure avoidance applied to validated results can be purpose-built
- 👍 Validated results can have proper error adjustments
- 👍 Validated results could be shared publicly to aid replication
- 👍 Synthetic data can be on full sample rather than a subsample
- 👍 Combined system could allow for sub-state analysis other than PUMA-level

Transparency is the way forward

- The Census Bureau is committed to openness in how it handles your data, and disclosure avoidance should be no different
 - Update on timeline and milestones related to disclosure avoidance modernization on our Disclosure Avoidance website:
<https://www.census.gov/programs-surveys/acs/methodology/disclosure-avoidance.html>
 - Gather use cases to understand how synthetic data, validation, and formal privacy can affect decisions
 - Conduct an external user test of the synthetic data and validation system
- Please send any questions on ACS disclosure avoidance to:
ACSprivacy@census.gov