

Improving Disclosure Avoidance Procedures for the Current Population Survey Public Use File

Federal Economic Statistics Advisory Committee (FESAC)
June 10, 2022

Eloise Parker, Assistant Director for Demographic Programs

Gary Benedetto, Assistant Center Chief, Center for Enterprise Dissemination – Disclosure Avoidance

Kyra Linse, Survey Director, Demographic Programs Directorate

U.S. Census Bureau

Context for Discussion

- Federal statistical agencies are bound by law to ensure the confidentiality of our survey respondents
 - Census Bureau products are protected by the Census Act (13 U.S. Code §§ 8(b) and 9)
 - Publications of all statistical agencies in the U.S. are protected under the Confidential Information Protection and Statistical Efficiency Act of 2018 (44 U.S. Code § 3563)
- Federal statistical products face increasingly disclosure risk
 - CPS PUF is no exception
- Census conducted a reidentification study on the Current Population Survey Public Use File (CPS PUF) in late 2020
 - Using a robust attacker file revealed vulnerabilities
 - Disclosure risk was most pronounced in lower-population geographies

Journey to Today

- Work ensued to apply additional protections to the 2022 PUF
- Announcement of changes posted in January 2022 prompted significant concerns within the user community
- Central concerns with that proposal:
 - Proposal would replace the public use file household identifier, precluding year-over-year time series analyses – and do so at a particularly sensitive time in the U.S. economy given the pandemic and associated stressors
 - Plans for rounding Usual Weekly Earnings and Usual Hourly Earnings Data would constrain analysis on wage growth and compromise input for tools like the Atlanta Federal Reserve Wage Tracker
 - Application of new methods would delay PUF release

Towards a Better Approach

- The Census Bureau acknowledged these user concerns
- Deferred planned changes to the 2022 PUF and released the file on the usual cycle, without delays
- The Bureau has researched alternatives for addressing potential disclosure avoidance issues to improve the balance between the obligation to protect confidentiality and the desire to meet users' needs
- Sharing that proposal today

New Proposal

- Established practice for CPS PUF is suppression of geographies with populations of less than 100,000
 - Goal to address vulnerabilities found in reidentification study for geographies with populations below 250,000
- To protect data between 100,000 and 250,000
 - Apply modeling solution to this level of geography only
 - Allows us to supply data at lower level instead of straight suppression
- Some rounding of wage and earnings
 - Less aggressive than earlier proposal

CBSA Size (Example) - Suppression

State	Internal		Population Size Category for Geo	Old PUF		Hypothetical PUF	
	Geocode	Met Status		CBSA	Met Status	CBSA	Met Status
AA	001	1 (Metro)	4 (500,000+)	001	1 (Metro)	001	1
AA	002	1 (Metro)	4 (500,000+)	002	1 (Metro)	002	1
AA	003	1 (Metro)	3 (250,000-499,999)	003	1 (Metro)	003	1
AA	004	1 (Metro)	1 (<100,000)	Other	1 (Metro)	Other	3
AA	005	1 (Metro)	2 (100,000-249,999)	005	1 (Metro)	Other	3
AA	006	1 (Metro)	2 (100,000-249,999)	006	1 (Metro)	Other	3
AA	007	1 (Metro)	1 (<100,000)	Other	1 (Metro)	Other	3
AA	008	2 (Non-met)	1 (<100,000)	Other	2 (Non-met)	Other	3
AA	009	2 (Non-met)	1 (<100,000)	Other	2 (Non-met)	Other	3

Synthesis Model

- Multinomial logistic regression predicting which “small” CBSA a household is in based on characteristics of the household and householder
- Separate regression model estimated for each state
- Only households within CBSA’s that are in the 100K-250K size category or in the “Other” category
- Took a data-driven approach to the model
 - Use many candidate explanatory variables at the household/householder level
 - Let the data determine the most important relationships using robust statistical methods

Example: Values of Dependent Variable

State	Internal		Population Size Category for Geo	Old PUF		Hypothetical PUF	
	Geocode	Met Status		CBSA	Met Status	CBSA	Met Status
AA	001	1 (Metro)	4 (500,000+)	001	1 (Metro)	001	1
AA	002	1 (Metro)	4 (500,000+)	002	1 (Metro)	002	1
AA	003	1 (Metro)	3 (250,000-499,999)	003	1 (Metro)	003	1
AA	004	1 (Metro)	1 (<100,000)	Other	1 (Metro)	Other	3
AA	005	1 (Metro)	2 (100,000-249,999)	005	1 (Metro)	Other	3
AA	006	1 (Metro)	2 (100,000-249,999)	006	1 (Metro)	Other	3
AA	007	1 (Metro)	1 (<100,000)	Other	1 (Metro)	Other	3
AA	008	2 (Non-met)	1 (<100,000)	Other	2 (Non-met)	Other	3
AA	009	2 (Non-met)	1 (<100,000)	Other	2 (Non-met)	Other	3

Illustration with Hypothetical Microdata

HH	ST	Internal CBSA	Old PUF CBSA	Met Stat	CBSA Size	HHR Age	HHR Earn
1	AA	006	006	1	2	52	2200
2	AA	006	006	1	2	35	600
3	AA	004	Other	1	1	64	1900
4	AA	005	005	1	2	43	1700
5	AA	008	Other	2	1	72	0
6	AA	003	003	1	3	38	800
7	AA	001	001	1	4	49	1200
8	AA	002	002	1	4	57	1000
9	AA	002	002	1	4	61	1500

Illustration with Hypothetical Microdata

HH	ST	Internal CBSA	Old PUF CBSA	Met Stat	CBSA Size	HHR Age	HHR Earn	Hypothetical Suppressed CBSA	Met Stat
1	AA	006	006	1	2	52	2200	Other	3
2	AA	006	006	1	2	35	600	Other	3
3	AA	004	Other	1	1	64	1900	Other	3
4	AA	005	005	1	2	43	1700	Other	3
5	AA	007	Other	2	1	72	0	Other	3
6	AA	003	003	1	3	38	800	003	1
7	AA	001	001	1	4	49	1200	001	1
8	AA	002	002	1	4	57	1000	002	1
9	AA	002	002	1	4	61	1500	002	1

Illustration with Hypothetical Microdata

HH	ST	Internal CBSA	Old PUF CBSA	Met Stat	CBSA Size	HHR Age	HHR Earn	Hypothetical Suppressed CBSA	Met Stat	Proposed Synthetic CBSA	Met Stat
1	AA	006	006	1	2	52	2200	Other	3	005	1
2	AA	006	006	1	2	35	600	Other	3	006	1
3	AA	004	Other	1	1	64	1900	Other	3	Other	1
4	AA	005	005	1	2	43	1700	Other	3	006	1
5	AA	007	Other	2	1	72	0	Other	3	Other	2
6	AA	003	003	1	3	38	800	003	1	003	1
7	AA	001	001	1	4	49	1200	001	1	001	1
8	AA	002	002	1	4	57	1000	002	1	002	1
9	AA	002	002	1	4	61	1500	002	1	002	1

Illustration with Hypothetical Microdata

HH	ST	Internal CBSA	Old PUF CBSA	Met Stat	CBSA Size	HHR Age	HHR Earn	Hypothetical Suppressed CBSA	Met Stat	Proposed Synthetic CBSA	Met Stat
1	AA	006	006	1	2	52	2200	Other	3	005	1
2	AA	006	006	1	2	35	600	Other	3	006	1
3	AA	004	Other	1	1	64	1900	Other	3	Other	1
4	AA	005	005	1	2	43	1700	Other	3	006	1
5	AA	007	Other	2	1	72	0	Other	3	Other	2
6	AA	003	003	1	3	38	800	003	1	003	1
7	AA	001	001	1	4	49	1200	001	1	001	1
8	AA	002	002	1	4	57	1000	002	1	002	1
9	AA	002	002	1	4	61	1500	002	1	002	1

Illustration with Hypothetical Microdata

HH	ST	Internal CBSA	Old PUF CBSA	Met Stat	CBSA Size	HHR Age	HHR Earn	Hypothetical Suppressed CBSA	Met Stat	Proposed Synthetic CBSA	Met Stat
1	AA	006	006	1	2	52	2200	Other	3	005	1
2	AA	006	006	1	2	35	600	Other	3	006	1
3	AA	004	Other	1	1	64	1900	Other	3	Other	1
4	AA	005	005	1	2	43	1700	Other	3	006	1
5	AA	007	Other	2	1	72	0	Other	3	Other	2
6	AA	003	003	1	3	38	800	003	1	003	1
7	AA	001	001	1	4	49	1200	001	1	001	1
8	AA	002	002	1	4	57	1000	002	1	002	1
9	AA	002	002	1	4	61	1500	002	1	002	1

Implications for Data Users

- What stays the same
 - All analyses at the state, census division, census region, and national level (including all analyses that do not condition on geography)
 - All analyses that condition on CBSA's that have a population of at least 250K.
- What changes
 - All analyses that condition on CBSA's that have a population below 250K.
 - Metropolitan vs non-metropolitan (in states where this distinction causes geos to fall below 250K)
- Alternative is suppression

Rounding of Wages

- Due to the synthesis of low-population geographies, wage and earnings rounding can be minimal—applied only to very-likely unique wage and earnings values.
- Dynamic topcode at top 3% reported
- Wages below the federal minimum wage will be flagged.
- Rounding proposal:

Hourly Wage Range	Rounding Amount
\$.01 - \$.07	Set to \$.05
\$.08 – \$19.99	\$.05
\$20.00 – \$39.99	\$.25
\$40.00 to topcode	\$.50

Implementation

- Phased approach
- No break in series
- Apply to 1st month-in-sample and continue
- Will take 16 months to phase in fully

Next Steps

- Looking for comments through July 15, 2022
- Final plan to be released around September 1, 2022
- Expected implementation with January data release -- February 3, 2023
- Send comments to : ADDP.CPS.PUF.List@census.gov