What Impact Do the 2020 Census Differential Privacy Methods Have on the National Survey on Drug Use and Health Sample?

Katie Morton, RTI International
Rong Cai, SAMHSA
Ahmed Khago, SAMHSA
Peilan Martin, RTI International
Matt Williams, RTI International

RTI INTERNATIONAL

Outline

- 2020 Census Differential Privacy Methods
- NSDUH Sample Design Summary
- NSDUH Sample Impact Evaluation
- Conclusion

2020 Census Differential Privacy Methods

- Disclosure avoidance methods are applied to every decennial Census product
- 2020 Census disclosure avoidance system was modernized to use differential privacy methods
  - Statistical noise added to every published estimate
  - No person or household can be identified with certainty
- Concern that differential privacy methods may impact ability to target demographic subgroups in survey samples
- U.S. Census Bureau released 2010 demonstration products that were treated with 2020 Census differential privacy methods
  - Used to evaluate the impact of the differential privacy methods on the NSDUH sample

2025-2027 NSDUH Sample Design

- Multi-stage area probability sample within each state and D.C.
  - Strata: State sampling regions
  - Stage 1: PSUs (census tracts)
  - Stage 2: SSUs (census block groups)
  - Stage 3: Dwelling units (households + eligible group quarters)
  - Stage 4: Individuals

2025-2027 NSDUH Sample Design (continued)

- Target 67,507 completed interviews annually
  - 25% 12 to 17
  - 25% 18 to 25
  - 15% 26 to 34
  - 20% 35 to 49
  - 15% 50+
- Oversample youth and young adults by:
  - Selecting PSUs and SSUs with probability proportional to composite measure of size (MOS) (defined on next slide)
  - Specifying within-household probabilities of selection by age group
  - Selecting 0, 1, or 2 people per household
- No oversampling of race/ethnicity groups

2025-2027 NSDUH Sample Design (continued)

- NSDUH composite measure of size (MOS)

$$MOS_{hk} = \sum_{d=1}^{5} f_h(d) C_{hk}(d)$$

Where:

- $f_h(d)$ is the sampling rate for age group d in state h
- $C_{hk}(d)$ is the 2020 Census population count for age group d in block k in state h

- **PSUs and SSUs with larger MOS have a greater chance of selection**

Impact of 2020 Census DPMs on NSDUH Area Samples

- Compute block-level MOS using both 2010 Census and 2010 demonstration data
- Aggregate MOS to census tract (PSU) and block group (SSU) levels
- Compare measures of size based on 2010 Census and 2010 demonstration data to determine impact on PSU and SSU samples
- Examine characteristics of census tracts and block groups with large absolute relative differences in MOS
  - absolute relative difference = $\frac{|MOS_{demo} - MOS_{2010}|}{MOS_{2010}}$
  - "large" = upper quartile of absolute relative differences among census tracts or block groups

Comparison of Census Tract and Block Group-Level Measures of Size

Census Tract-Level MOS

Data Source	Minimum	Maximum	Mean	Rho
2010 Census	0	35.46114	0.92798	0.99998*
Demo Data	0	35.33252	0.92798	

Block Group-Level MOS

Data Source	Minimum	Maximum	Mean	Rho
2010 Census	0	30.64562	0.31080	0.99927*
Demo Data	0	30.57174	0.31080	

*$p<0.001$

Characteristics of Census Tracts and Block Groups with "Large" Absolute Relative Differences in MOS

	Census Tracts		Census Block Groups	
Data Source	**Large MOS Difference**	**Remainder**	**Large MOS Difference**	**Remainder**
2010 Census				
Minimum	0.00	0.00	0.00	0.00
Mean	**0.71**	**1.00**	**0.23**	**0.34**
Maximum	10.36	35.46	6.06	30.65
Demo Data				
Minimum	0.00	0.00	0.00	0.00
Mean	**0.71**	**1.00**	**0.23**	**0.34**
Maximum	10.41	35.33	5.80	30.57

Characteristics of Census Tracts and Block Groups with "Large" Absolute Relative Differences in MOS (continued)

	Census Tracts		Census Block Groups	
Characteristic	Large MOS Difference	Remainder	Large MOS Difference	Remainder
Percent Urban	16.11%	9.09%*	10.28%	15.40%*
Percent 12-17	9.22%	9.70%*	9.37%	9.70%*
Percent AIAN	1.25%	0.90%*	0.93%	0.99%*

AIAN = American Indian or Alaska Native

*Estimate is significantly different from Large MOS Difference estimate at the 0.05 level.

Impact of 2020 Census DPMs on Sample Yields

- Simulate sample yields for the 2023 NSDUH using both 2010 Census and 2010 demonstration data
  - For each domain, compute expected people per dwelling unit (DU) as the domain population count (2010 Census or demonstration data) divided by occupied DUs in the sampled area
  - Multiply sample DUs by expected people per DU and person probability of selection
  - Adjust for DU eligibility, DU nonresponse, maximum of 2 selected people per DU, and person-level nonresponse
- Scale simulated sample yields to 67,507 completed interviews
- Compare expected sample yields by age group (targeted) and race/ethnicity (not targeted)

Expected Sample Yields Using 2010 Census and Demonstration Data

	2010 Census		Demo Data	
AGE GROUP	**Count**	**Percent**	**Count**	**Percent**
Total	67,507	100.00%	67,507	100.00%
12-17	16,883	25.01%	16,893	25.02%
18-25	**18,907**	**28.01%**	**18,860**	**27.94%**
36-34	9,660	14.31%	9,647	14.29%
35-49	13,925	20.63%	13,948	20.66%
50+	8,132	12.05%	8,159	12.09%

Bold = largest difference

Expected Sample Yields Using 2010 Census and Demonstration Data

	2010 Census		Demo Data	
RACE/ETHNICITY	**Count**	**Percent**	**Count**	**Percent**
Total	67,507	100.00%	67,507	100.00%
Hispanic or Latino	10,240	15.17%	10,223	15.14%
White	**43,086**	**63.82%**	**43,109**	**63.86%**
Black	8,348	12.37%	8,368	12.40%
AIAN	866	1.28%	861	1.28%
Asian	3,076	4.56%	3,069	4.55%
NHOPI	188	0.28%	189	0.28%
Other	1,703	2.52%	1,688	2.50%

AIAN = American Indian or Alaska Native; NHOPI = Native Hawaiian or Other Pacific Islander.

Bold = largest difference

Conclusions

- The 2020 Census differential privacy methods are not expected to have a significant impact on the NSDUH sample under the current design.
- Composite measures of size based on 5 age groups are very similar using the treated and untreated data.
- PSUs and SSUs that are most impacted by the data perturbation methods are less likely to be selected.
- The treated data is expected to yield a similar sample by age and race/ethnicity as the untreated data.

Thank you

Contact: Katie Morton| email: bowman@rti.org