Sampling with Administrative Records in the National Survey of Children's Health

Scott Albrecht
US Census Bureau

Jason Fields
US Census Bureau

Keith Finlay
US Census Bureau

March 9, 2018

United States™ Census Bureau

Disclaimer: This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the US Census Bureau.

National Survey of Children's Health

- Sponsored by the Maternal and Child Health Bureau of the Health Resources and Services Administrative (HHS)
- Census Bureau became the collection agent in 2015
  - Transition from phone frame to address frame
- Population is households with children
  - Sampled housing units receive screener for child presence
  - One child from positive screeners chosen for topical survey
  - State-level representation ($\approx$ equal sample yields)
- The Census Bureau used administrative records to target households likely to have children
  - Sampling costs were half what they would be under simple random sampling

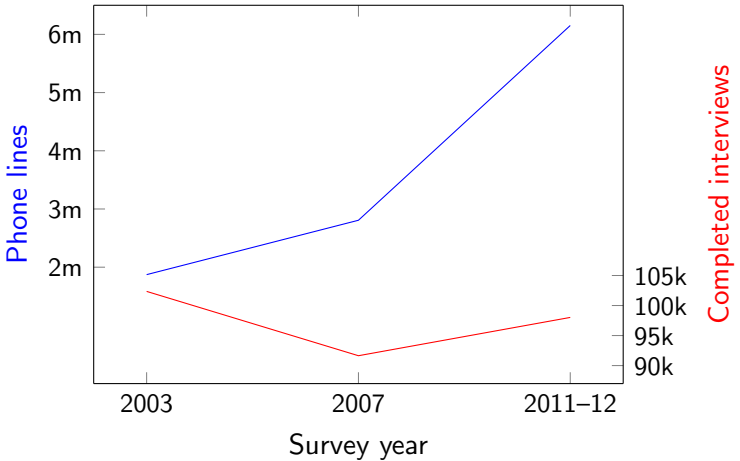
National Survey of Children's Health

- Sponsored by the Maternal and Child Health Bureau of the Health Resources and Services Administrative (HHS)
- Census Bureau became the collection agent in 2015
  - Transition from phone frame to address frame
- Population is households with children
  - Sampled housing units receive screener for child presence
  - One child from positive screeners chosen for topical survey
  - State-level representation ($\approx$ equal sample yields)
- The Census Bureau used administrative records to target households likely to have children
  - Sampling costs were half what they would be under simple random sampling

National Survey of Children's Health

- Sponsored by the Maternal and Child Health Bureau of the Health Resources and Services Administrative (HHS)
- Census Bureau became the collection agent in 2015
  - Transition from phone frame to address frame
- Population is households with children
  - Sampled housing units receive screener for child presence
  - One child from positive screeners chosen for topical survey
  - State-level representation ($\approx$ equal sample yields)
- The Census Bureau used administrative records to target households likely to have children
  - Sampling costs were half what they would be under simple random sampling

Phone frames were increasingly inefficient



Source: Blumberg et al. (2005), Blumberg et al. (2012), and Bramlett et al. (2017)

United States Census Bureau

Identify strata from administrative records

- Identify households likely to have children
- Oversample from those households to reduce costs
- Two approaches
  - Rule based with explicit links from children to addresses
  - Optimized strata based on probabilistic measures of child presence under a coverage constraint

Year	Strata definitions	
2016	Explicit links Stratum 1	No explicit links Stratum 2
2017	Explicit links Stratum 1	Child likely Stratum 2a	Child unlikely Stratum 2b

Identify strata from administrative records

- Identify households likely to have children
- Oversample from those households to reduce costs
- Two approaches
  - **Rule-based** with explicit links from children to addresses
  - Optimized strata based on **probabilistic** measures of child presence under a coverage constraint

Year	Strata definitions		
2016	Explicit links Stratum 1	No explicit links Stratum 2	
2017	Explicit links Stratum 1	Child likely Stratum 2a	Child unlikely Stratum 2b

Identify strata from administrative records

- Identify households likely to have children
- Oversample from those households to reduce costs
- Two approaches
  - **Rule-based** with explicit links from children to addresses
  - Optimized strata based on **probabilistic** measures of child presence under a coverage constraint

Year	Strata definitions		
2016	Explicit links Stratum 1	No explicit links Stratum 2	
2017	Explicit links Stratum 1	Child likely Stratum 2a	Child unlikely Stratum 2b

United States Census Bureau

Administrative records used for rule-based flag

- Registers: SSA Numident, Census Master Address File
- Tax data: 1040s, 1099s
- Federal programs: Medicare, Indian Health Service, HUD housing assistance, Selective Service, Postal Service National Change of Address
- 2010 Decennial Census data
- These files are used to create population-level auxiliary files that link:
  - Children to parents
  - Children and parents to addresses
- Child-present flag is applied to the Master Address File before sampling

Administrative records used for rule-based flag

- Registers: SSA Numident, Census Master Address File
- Tax data: 1040s, 1099s
- Federal programs: Medicare, Indian Health Service, HUD housing assistance, Selective Service, Postal Service National Change of Address
- 2010 Decennial Census data
- These files are used to create population-level auxiliary files that link:
  - Children to parents
  - Children and parents to addresses
- Child-present flag is applied to the Master Address File before sampling

Administrative records used for rule-based flag

- Registers: SSA Numident, Census Master Address File
- Tax data: 1040s, 1099s
- Federal programs: Medicare, Indian Health Service, HUD housing assistance, Selective Service, Postal Service National Change of Address
- 2010 Decennial Census data
- These files are used to create population-level auxiliary files that link:
  - Children to parents
  - Children and parents to addresses
- Child-present flag is applied to the Master Address File before sampling

2016 rule-based frame performed well in ACS audit

- Any children in household, 2016 NSCH child flag versus 2014 ACS response

	Observed ACS households		
NSCH child flags	No children	Any children	Total
No children	92.2%	7.8%	100.0%
Any children	25.2%	74.8%	100.0%
Total	74.6%	25.4%	100.0%
N (ACS households)	2,322,722		

Source: authors' calculations.

2016 rule-based frame performed well in ACS audit

- Any children in household, 2016 NSCH child flag versus 2014 ACS response

	Observed ACS households		
NSCH child flags	No children	Any children	Total
No children	92.2%	7.8%	100.0%
Any children	25.2%	74.8%	100.0%
Total	74.6%	25.4%	100.0%
N (ACS households)	2,322,722		

Source: authors' calculations.

2016 frame performed well in sampling response

- Flagged households 9 times more likely to report children: 77.3% versus 8.5%
- Small characteristic differences across strata

Respondent characteristics	Stratum 1	Stratum 2
Child		
Hispanic	24.6%	24.5%
White alone	68.4%	63.3%
Foreign born	3.3%	7.7%
Reference person		
Education > HS	67.1%	64.1%

Source: authors' calculations.

2016 frame performed well in sampling response

- Flagged households 9 times more likely to report children: 77.3% versus 8.5%
- Small characteristic differences across strata

Respondent characteristics	Stratum 1	Stratum 2
Child		
Hispanic	24.6%	24.5%
White alone	68.4%	63.3%
Foreign born	3.3%	7.7%
Reference person		
Education > HS	67.1%	64.1%

Source: authors' calculations.

Absence of selection where we expect it

2016 frame performed well in sampling response

- Flagged households 9 times more likely to report children: 77.3% versus 8.5%
- Small characteristic differences across strata

Respondent characteristics	Stratum 1	Stratum 2
Child		
Hispanic	24.6%	24.5%
White alone	68.4%	63.3%
Foreign born	3.3%	7.7%
Reference person		
Education > HS	67.1%	64.1%

Source: authors' calculations.

Moderate selection in the direction we expect

2016 frame performed well in sampling response

- Flagged households 9 times more likely to report children: 77.3% versus 8.5%
- Small characteristic differences across strata

Respondent characteristics	Stratum 1	Stratum 2
Child		
Hispanic	24.6%	24.5%
White alone	68.4%	63.3%
Foreign born	3.3%	7.7%
Reference person		
Education > HS	67.1%	64.1%

Source: authors' calculations.

Numident selection + adrec lag for recent movers

Other 2016 sample features

- Stratum 2 contains relatively more very young children
  - 14% of completed topicals from Stratum 2 are less than 1 year old (versus 5% under a uniform age distribution)
  - Reflects administrative records lag and has implications for all surveys
  - Even with timely administrative records, ACS test data are 1–2 years old
- Differential costs in 2016 sampling
  - Children are present in only 7–8% of households in Stratum 2
  - Cost of collecting information about those children is incredibly high ($536 per topical versus $68 per topical in Stratum 1)
  - Desire to further increase sampling efficiency for 2017 NSCH

Other 2016 sample features

- Stratum 2 contains relatively more very young children
  - 14% of completed topicals from Stratum 2 are less than 1 year old (versus 5% under a uniform age distribution)
  - Reflects administrative records lag and has implications for all surveys
  - Even with timely administrative records, ACS test data are 1–2 years old
- Differential costs in 2016 sampling
  - Children are present in only 7–8% of households in Stratum 2
  - Cost of collecting information about those children is incredibly high ($536 per topical versus $68 per topical in Stratum 1)
  - Desire to further increase sampling efficiency for 2017 NSCH

Added probabilistic strata to 2017 NSCH frame

- Split Stratum 2 (those without explicit links) into
  - 2a. Households likely to have children conditional on administrative records
  - 2b. Households unlikely to have children, who will not be sampled
- Problem:
  - Need data on child presence in a sample
  - Predictive administrative data available from population
  - Estimate a model on ACS sample
  - Predict to population of addresses
  - Optimize strata with predictions and a coverage constraint

Added probabilistic strata to 2017 NSCH frame

- Split Stratum 2 (those without explicit links) into
  - 2a. Households likely to have children conditional on administrative records
  - 2b. Households unlikely to have children, who will not be sampled
- Problem:
  - Need data on child presence in a sample
  - Predictive administrative data available from population
  - Estimate a model on ACS sample
  - Predict to population of addresses
  - Optimize strata with predictions and a coverage constraint

Modeling and optimizing probabilistic strata

- Used a linear probability model of child-presence
  - Estimated at the address-level using the 2015 ACS sample
  - Predictions estimated for the Master Address File population
- Predictors:
  - Household age structure synthesized from administrative records
  - Child indicators synthesized from administrative records
  - Various commercial measures
  - Missingness indicators for family structure, commercial data
  - Proportion of residents of block group who are children
- Optimization by state
  - Minimize the size of Stratum 2a while maintaining coverage of at least 95% of households with children for each state

Modeling and optimizing probabilistic strata

- Used a linear probability model of child-presence
  - Estimated at the address-level using the 2015 ACS sample
  - Predictions estimated for the Master Address File population
- Predictors:
  - Household age structure synthesized from administrative records
  - Child indicators synthesized from administrative records
  - Various commercial measures
  - Missingness indicators for family structure, commercial data
  - Proportion of residents of block group who are children
- Optimization by state
  - Minimize the size of Stratum 2a while maintaining coverage of at least 95% of households with children for each state

Modeling and optimizing probabilistic strata

- Used a linear probability model of child-presence
  - Estimated at the address-level using the 2015 ACS sample
  - Predictions estimated for the Master Address File population
- Predictors:
  - Household age structure synthesized from administrative records
  - Child indicators synthesized from administrative records
  - Various commercial measures
  - Missingness indicators for family structure, commercial data
  - Proportion of residents of block group who are children
- Optimization by state
  - Minimize the size of Stratum 2a while maintaining coverage of at least 95% of households with children for each state

Optimal strata for the 2017 NSCH sample frame

- Optimal 2017 strata applied to 2015 ACS data

		Strata distribution			Child-present rates			Coverage
	N	$p(S_1)$	$p(S_{2a})$	$p(S_{2b})$	$p(C\|S_1)$	$p(C\|S_{2a})$	$p(C\|S_{2b})$	$p(S_1 \cup S_{2a}\|C)$
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96

Source: authors' calculations.

- Further validation with preliminary 2016 NSCH and ACS responses to assess overfitting
- Small characteristic imbalances:
  - Poorer, rural, renting households headed by non-citizens and minorities with no health insurance who have recently moved
  - Future: incorporate representativeness constraints

Optimal strata for the 2017 NSCH sample frame

- Optimal 2017 strata applied to 2015 ACS data

		Strata distribution			Child-present rates			Coverage
	N	$p(S_1)$	$p(S_{2a})$	$p(S_{2b})$	$p(C\|S_1)$	$p(C\|S_{2a})$	$p(C\|S_{2b})$	$p(S_1 \cup S_{2a}\|C)$
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96

Source: authors' calculations.

- Further validation with preliminary 2016 NSCH and ACS responses to assess overfitting
- Small characteristic imbalances:
  - Poorer, rural, renting households headed by non-citizens and minorities with no health insurance who have recently moved
  - Future: incorporate representativeness constraints

Optimal strata for the 2017 NSCH sample frame

- Optimal 2017 strata applied to 2015 ACS data

		Strata distribution			Child-present rates			Coverage
	N	$p(S_1)$	$p(S_{2a})$	$p(S_{2b})$	$p(C\|S_1)$	$p(C\|S_{2a})$	$p(C\|S_{2b})$	$p(S_1 \cup S_{2a}\|C)$
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96

Source: authors' calculations.

- Further validation with preliminary 2016 NSCH and ACS responses to assess overfitting
- Small characteristic imbalances:
  - Poorer, rural, renting households headed by non-citizens and minorities with no health insurance who have recently moved
  - Future: incorporate representativeness constraints

Optimal strata for the 2017 NSCH sample frame

- Optimal 2017 strata applied to 2015 ACS data

		Strata distribution			Child-present rates			Coverage
	N	$p(S_1)$	$p(S_{2a})$	$p(S_{2b})$	$p(C\|S_1)$	$p(C\|S_{2a})$	$p(C\|S_{2b})$	$p(S_1 \cup S_{2a}\|C)$
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96

Source: authors' calculations.

- Further validation with preliminary 2016 NSCH and ACS responses to assess overfitting
- Small characteristic imbalances:
  - Poorer, rural, renting households headed by non-citizens and minorities with no health insurance who have recently moved
  - Future: incorporate representativeness constraints

Optimal strata for the 2017 NSCH sample frame

- Optimal 2017 strata applied to 2015 ACS data

		Strata distribution			Child-present rates			Coverage
	N	$p(S_1)$	$p(S_{2a})$	$p(S_{2b})$	$p(C\|S_1)$	$p(C\|S_{2a})$	$p(C\|S_{2b})$	$p(S_1 \cup S_{2a}\|C)$
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96

Source: authors' calculations.

- Further validation with preliminary 2016 NSCH and ACS responses to assess overfitting
- Small characteristic imbalances:
  - Poorer, rural, renting households headed by non-citizens and minorities with no health insurance who have recently moved
  - Future: incorporate representativeness constraints

Optimal strata for the 2017 NSCH sample frame

- Optimal 2017 strata applied to 2015 ACS data

		Strata distribution			Child-present rates			Coverage
	N	$p(S_1)$	$p(S_{2a})$	$p(S_{2b})$	$p(C\|S_1)$	$p(C\|S_{2a})$	$p(C\|S_{2b})$	$p(S_1 \cup S_{2a}\|C)$
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96

Source: authors' calculations.

- Further validation with preliminary 2016 NSCH and ACS responses to assess overfitting
- Small characteristic imbalances:
  - Poorer, rural, renting households headed by non-citizens and minorities with no health insurance who have recently moved
  - Future: incorporate representativeness constraints

Preliminary 2017 sampling results

- 2017 NSCH sample period ended February 16, 2018
- Caveat: smaller states overweighted and oversample varies by state

Stratum	2015 ACS	Preliminary 2017 NSCH screeners
1. Explicit links	78%	77%
2a. Not explicit, but likely	17%	13%
2b. Not explicit, but unlikely	3%	not sampled
N (households)	2,208,198	58,653

Source: authors' calculations.

Preliminary 2017 sampling results

- 2017 NSCH sample period ended February 16, 2018
- Caveat: smaller states overweighted and oversample varies by state

Stratum	2015 ACS	Preliminary 2017 NSCH screeners
1. Explicit links	78%	77%
2a. Not explicit, but likely	17%	13%
2b. Not explicit, but unlikely	3%	not sampled
N (households)	2,208,198	58,653

Source: authors' calculations.

Preliminary 2017 sampling results

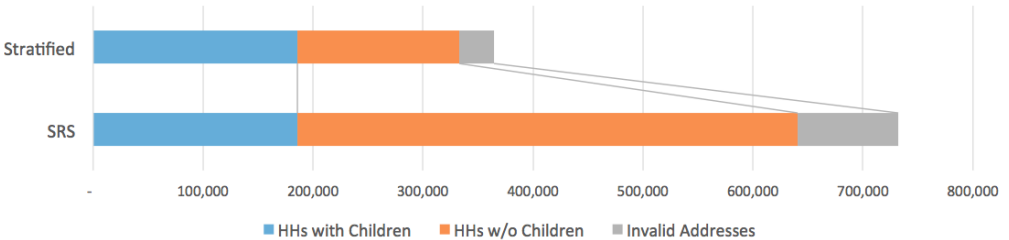
- 2017 NSCH sample period ended February 16, 2018
- Caveat: smaller states overweighted and oversample varies by state

Stratum	2015 ACS	Preliminary 2017 NSCH screeners
1. Explicit links	78%	77%
2a. Not explicit, but likely	17%	13%
2b. Not explicit, but unlikely	3%	not sampled
N (households)	2,208,198	58,653

Source: authors' calculations.

Stratified sampling increases efficiency and reduces costs

- Required sample sizes under stratified sampling with flag oversample versus simple random sampling, 2016 NSCH



Source: authors' calculations.

Stratified sampling improves precision

- Standard errors of estimated respondent characteristics under stratified oversampling versus simple random sampling, 2016 NSCH respondents

Characteristic/statistic	Stratified oversampling	Simple random sampling	Delta
Total kids			
Mean	0.011	0.014	-26.4%
Federal poverty ratio			
Mean	1.396	1.871	-25.4%
Health insurance			
Yes	0.003	0.004	-23.3%

Source: authors' calculations.

Risks of stratified sampling

- Nonrandom sampling
  - Record linkage selection bias
  - Assess how observables vary across strata in the ACS
  - Frames from commercial vendors have greater sample bias
    - In 2015 NSCH pre-test, 66.9% of respondents had completed BA versus 41.3% in 2013 ACS
    - Median household income was higher for every education group
- Undercoverage of children (with unsampled Stratum 2b)
  - 10% of children in the Numident were not linked to any parent
  - 13% of children in the Numident were not linked to any address
- Trade off these risks against benefits of increased sample yields
- Improve communication of more complex frame to sponsor and data users

Risks of stratified sampling

- Nonrandom sampling
  - Record linkage selection bias
  - Assess how observables vary across strata in the ACS
  - Frames from commercial vendors have greater sample bias
    - In 2015 NSCH pre-test, 66.9% of respondents had completed BA versus 41.3% in 2013 ACS
    - Median household income was higher for every education group
- Undercoverage of children (with unsampled Stratum 2b)
  - 10% of children in the Numident were not linked to any parent
  - 13% of children in the Numident were not linked to any address
- Trade off these risks against benefits of increased sample yields
- Improve communication of more complex frame to sponsor and data users

Risks of stratified sampling

- Nonrandom sampling
  - Record linkage selection bias
  - Assess how observables vary across strata in the ACS
  - Frames from commercial vendors have greater sample bias
    - In 2015 NSCH pre-test, 66.9% of respondents had completed BA versus 41.3% in 2013 ACS
    - Median household income was higher for every education group
- Undercoverage of children (with unsampled Stratum 2b)
  - 10% of children in the Numident were not linked to any parent
  - 13% of children in the Numident were not linked to any address
- Trade off these risks against benefits of increased sample yields
- Improve communication of more complex frame to sponsor and data users

Risks of stratified sampling

- Nonrandom sampling
  - Record linkage selection bias
  - Assess how observables vary across strata in the ACS
  - Frames from commercial vendors have greater sample bias
    - In 2015 NSCH pre-test, 66.9% of respondents had completed BA versus 41.3% in 2013 ACS
    - Median household income was higher for every education group
- Undercoverage of children (with unsampled Stratum 2b)
  - 10% of children in the Numident were not linked to any parent
  - 13% of children in the Numident were not linked to any address
- Trade off these risks against benefits of increased sample yields
- Improve communication of more complex frame to sponsor and data users

Risks of stratified sampling

- Nonrandom sampling
  - Record linkage selection bias
  - Assess how observables vary across strata in the ACS
  - Frames from commercial vendors have greater sample bias
    - In 2015 NSCH pre-test, 66.9% of respondents had completed BA versus 41.3% in 2013 ACS
    - Median household income was higher for every education group
- Undercoverage of children (with unsampled Stratum 2b)
  - 10% of children in the Numident were not linked to any parent
  - 13% of children in the Numident were not linked to any address
- Trade off these risks against benefits of increased sample yields
- Improve communication of more complex frame to sponsor and data users

Conclusion

- Administrative records are an important tool for improving coverage and reducing costs of household surveys
- Probabilistic techniques can further increase sampling efficiency but with diminishing returns
- ACS gives us a nice reference point for assessing use of administrative records
- Need more research on administrative records selection

Conclusion

- Administrative records are an important tool for improving coverage and reducing costs of household surveys
- Probabilistic techniques can further increase sampling efficiency but with diminishing returns
- ACS gives us a nice reference point for assessing use of administrative records
- Need more research on administrative records selection

Conclusion

- Administrative records are an important tool for improving coverage and reducing costs of household surveys
- Probabilistic techniques can further increase sampling efficiency but with diminishing returns
- ACS gives us a nice reference point for assessing use of administrative records
- Need more research on administrative records selection

Conclusion

- Administrative records are an important tool for improving coverage and reducing costs of household surveys
- Probabilistic techniques can further increase sampling efficiency but with diminishing returns
- ACS gives us a nice reference point for assessing use of administrative records
- Need more research on administrative records selection

References I

Blumberg, Stephen J, Erin B Foster, Alicia M Frasier, Jennifer Satorius, Ben J Skalland, Kari L Nysse-Carris, Heather M Morrison, Sadeq R Chowdhury, and Kathleen S O'Connor. 2012. Design and Operation of the National Survey of Children's Health, 2007. *Vital and Health Statistics* 1 (55): 1–149. https://www.ncbi.nlm.nih.gov/pubmed/22834229.

Blumberg, Stephen J, Lorayn Olson, Martin R Frankel, Larry Osborn, K P Srinath, and Pamela Giambo. 2005. Design and Operation of the National Survey of Children's Health, 2003. *Vital and Health Statistics* 1 (43): 1–131. https://www.ncbi.nlm.nih.gov/pubmed/25078122.

Bramlett, Matthew D, Stephen J Blumberg, Benjamin Zablotsky, Jacquelyn M George, A Elizabeth Ormson, Alicia M Frasier, Danielle M Vstecka, et al. 2017. Design and Operation of the National Survey of Children's Health, 2011–2012. *Vital and Health Statistics* 1 (59): 1–256. https://www.ncbi.nlm.nih.gov/pubmed/28796596.

US Census Bureau. 2017. 2016 National Survey of Children's Health: Source and Accuracy Statement. https://www.census.gov/content/dam/Census/programs-surveys/nsch/tech-documentation/source-accuracy-statement/NSCH%202016%20Source%20and%20Accuracy%20Statement%20v2.pdf.

2016 NSCH strata and oversample rates, 2014 ACS

State	Oversample rate	$p(C)$	N	N_{S_1}	N_{S_2}
Alabama	5.4	23.0%	7,827	4,636	3,191
Alaska	3.8	26.8%	8,468	4,099	4,369
Arizona	5.4	23.7%	7,553	4,486	3,067
Arkansas	4.9	24.1%	7,965	4,576	3,389
California	5.3	29.8%	6,407	4,264	2,142
Colorado	5.8	26.5%	6,549	4,122	2,427
Connecticut	6.7	24.8%	6,526	4,361	2,164
Delaware	6.8	23.1%	6,890	4,552	2,338
District of Columbia	7.2	17.1%	8,439	4,970	3,468
Florida	6.5	19.9%	8,035	4,938	3,097
Georgia	4.7	27.7%	7,238	4,469	2,769
Hawaii	3.4	26.4%	9,490	3,524	5,966
Idaho	5.4	26.6%	6,812	4,205	2,607
Illinois	6.3	25.8%	6,573	4,428	2,146
Indiana	6.4	25.6%	6,542	4,425	2,117
Iowa	7.9	25.2%	5,938	4,238	1,700

2016 NSCH strata and oversample rates, 2014 ACS

State	Oversample rate	$p(C)$	N	N_{S_1}	N_{S_2}
Kansas	6.7	26.5%	6,318	4,401	1,918
Kentucky	5.2	25.1%	7,380	4,410	2,969
Louisiana	4.7	24.9%	8,057	4,841	3,216
Maine	8.3	17.4%	7,385	4,484	2,901
Maryland	6.1	26.6%	6,511	4,363	2,148
Massachusetts	7.3	24.5%	6,237	4,253	1,984
Michigan	8.6	22.7%	6,205	4,423	1,782
Minnesota	8.3	24.8%	5,871	4,217	1,654
Mississippi	4.6	25.9%	7,856	4,663	3,193
Missouri	6.4	23.6%	6,940	4,512	2,427
Montana	6.4	20.4%	7,605	4,352	3,253
Nebraska	6.8	26.6%	6,071	4,147	1,925
Nevada	4.8	24.1%	8,031	4,466	3,565
New Hampshire	7.9	21.6%	6,460	4,243	2,217
New Jersey	6.0	26.9%	6,447	4,247	2,200
New Mexico	4.4	23.8%	8,582	4,345	4,237
New York	4.8	24.2%	7,918	4,307	3,612
North Carolina	5.6	24.4%	7,215	4,448	2,766

United States Census Bureau

2016 NSCH strata and oversample rates, 2014 ACS

State	Oversample rate	$p(C)$	N	N_{S_1}	N_{S_2}
North Dakota	5.8	24.1%	7,038	4,151	2,887
Ohio	7.5	24.2%	6,402	4,497	1,905
Oklahoma	4.5	25.8%	7,908	4,490	3,418
Oregon	7.0	24.0%	6,414	4,224	2,190
Pennsylvania	7.7	22.7%	6,461	4,391	2,070
Rhode Island	6.8	23.3%	6,713	4,359	2,354
South Carolina	6.0	22.5%	7,543	4,694	2,849
South Dakota	5.5	23.7%	7,441	4,381	3,059
Tennessee	5.4	25.1%	7,241	4,500	2,741
Texas	4.1	30.7%	7,132	4,254	2,878
Utah	4.4	35.2%	6,126	4,081	2,045
Vermont	7.1	19.8%	7,151	4,105	3,045
Virginia	6.0	26.8%	6,513	4,314	2,199
Washington	6.2	25.4%	6,583	4,269	2,314
West Virginia	5.0	20.3%	9,116	4,165	4,952
Wisconsin	8.0	23.7%	6,136	4,267	1,869
Wyoming	4.8	24.1%	7,894	4,193	3,701

2017 NSCH strata, 2015 ACS, valid addresses audit

State	N	$p(S1)$	$p(S2)$	$p(S3)$	$p(C \mid S1)$	$p(C \mid S2)$	$p(C \mid S3)$	$p(!S3 \mid C)$
US	2,208,198	0.24	0.39	0.37	0.77	0.17	0.03	0.96
AL	35,869	0.23	0.43	0.34	0.74	0.15	0.03	0.97
AK	6,485	0.21	0.79	0.00	0.73	0.14	0.00	1.00
AZ	41,371	0.24	0.43	0.33	0.74	0.16	0.02	0.98
AR	21,346	0.23	0.47	0.30	0.75	0.16	0.03	0.97
CA	212,245	0.28	0.32	0.40	0.80	0.22	0.04	0.96
CO	37,090	0.24	0.37	0.39	0.81	0.18	0.03	0.96
CT	22,739	0.23	0.36	0.41	0.78	0.17	0.03	0.96
DE	6,888	0.22	0.33	0.45	0.76	0.15	0.02	0.97
DC	4,625	0.20	0.39	0.41	0.71	0.12	0.02	0.96
FL	118,291	0.21	0.34	0.44	0.70	0.15	0.02	0.96
GA	54,608	0.27	0.38	0.35	0.75	0.19	0.04	0.96
HI	9,347	0.18	0.58	0.24	0.73	0.24	0.03	0.98
ID	10,904	0.23	0.41	0.35	0.78	0.17	0.03	0.97
IL	92,002	0.25	0.39	0.36	0.78	0.17	0.03	0.97
IN	45,984	0.25	0.37	0.39	0.77	0.17	0.03	0.97
IA	28,792	0.24	0.76	0.00	0.80	0.08	0.00	1.00

United States Census Bureau

2017 NSCH strata, 2015 ACS, valid addresses audit

State	N	$p(S1)$	$p(S2)$	$p(S3)$	$p(C \mid S1)$	$p(C \mid S2)$	$p(C \mid S3)$	$p(!S3 \mid C)$
KS	24,639	0.25	0.41	0.34	0.78	0.15	0.02	0.97
KY	32,190	0.24	0.46	0.30	0.76	0.16	0.03	0.97
LA	29,836	0.25	0.41	0.34	0.70	0.17	0.04	0.96
ME	15,488	0.16	0.54	0.30	0.79	0.09	0.02	0.97
MD	37,925	0.27	0.34	0.39	0.79	0.18	0.03	0.96
MA	41,968	0.23	0.35	0.42	0.80	0.17	0.03	0.96
MI	95,207	0.22	0.33	0.45	0.79	0.15	0.03	0.96
MN	69,076	0.23	0.35	0.42	0.83	0.15	0.02	0.97
MS	17,539	0.26	0.46	0.28	0.70	0.17	0.03	0.97
MO	47,573	0.24	0.41	0.36	0.76	0.15	0.03	0.97
MT	10,044	0.18	0.68	0.14	0.77	0.11	0.01	0.99
NE	18,292	0.25	0.60	0.16	0.82	0.10	0.02	0.99
NV	17,793	0.24	0.44	0.32	0.71	0.16	0.03	0.97
NH	10,754	0.19	0.44	0.37	0.80	0.12	0.02	0.97
NJ	54,999	0.26	0.34	0.40	0.80	0.20	0.03	0.96
NM	14,112	0.20	0.61	0.19	0.69	0.16	0.02	0.98
NY	133,044	0.23	0.44	0.33	0.75	0.17	0.03	0.97
NC	66,516	0.23	0.39	0.38	0.76	0.16	0.03	0.96

United States Census Bureau

2017 NSCH strata, 2015 ACS, valid addresses audit

State	N	$p(S1)$	$p(S2)$	$p(S3)$	$p(C \mid S1)$	$p(C \mid S2)$	$p(C \mid S3)$	$p(!S3 \mid C)$
ND	8,518	0.21	0.63	0.16	0.77	0.10	0.02	0.99
OH	87,385	0.24	0.35	0.42	0.79	0.16	0.03	0.96
OK	41,629	0.25	0.55	0.20	0.73	0.15	0.03	0.98
OR	25,223	0.23	0.40	0.37	0.80	0.15	0.02	0.97
PA	113,690	0.22	0.35	0.43	0.80	0.15	0.03	0.96
RI	6,517	0.21	0.37	0.42	0.79	0.17	0.02	0.96
SC	31,694	0.23	0.39	0.38	0.72	0.15	0.03	0.96
SD	9,130	0.23	0.57	0.21	0.80	0.12	0.02	0.98
TN	42,585	0.25	0.39	0.37	0.76	0.16	0.03	0.96
TX	142,040	0.29	0.38	0.34	0.76	0.21	0.04	0.96
UT	17,599	0.34	0.34	0.32	0.83	0.21	0.04	0.96
VT	8,384	0.16	0.78	0.06	0.80	0.08	0.02	0.99
VA	52,723	0.26	0.32	0.42	0.80	0.17	0.03	0.95
WA	46,227	0.25	0.40	0.35	0.80	0.16	0.03	0.96
WV	13,003	0.20	0.80	0.00	0.74	0.11	0.00	1.00
WI	72,065	0.22	0.31	0.48	0.82	0.16	0.03	0.96
WY	4,205	0.20	0.46	0.34	0.75	0.18	0.03	0.97

2017 NSCH strata and oversample rates, 2015 ACS

State	Oversample rate	BW_S1	BW_S2a
Alabama	2.38	215.11	511.06
Alaska	1.33	37.44	49.84
Arizona	2.40	337.22	810
Arkansas	2.27	149.37	338.37
California	2.67	2062.61	5512.64
Colorado	2.40	354.05	848.36
Connecticut	2.82	205.58	579.05
Delaware	3.68	48.76	179.41
DC	4.31	26.23	113.09
Florida	3.47	858.68	2978.27
Georgia	2.56	523.77	1342.26
Hawaii	1.63	71.23	116.32
Idaho	1.93	115.79	223.67
Illinois	2.75	797.94	2191.24
Indiana	2.40	384.51	922.49
Iowa	4.06	209.04	848.65

2017 NSCH strata and oversample rates, 2015 ACS

State	Oversample rate	BW_S1	BW_S2a
Kansas	3.12	175.73	548.81
Kentucky	3.13	239.5	748.57
Louisiana	2.39	197.49	471.53
Maine	2.88	67.97	196.09
Maryland	2.88	394.83	1138.72
Massachusetts	2.81	432.18	1216.22
Michigan	3.44	631.75	2174.99
Minnesota	3.20	433.07	1384.3
Mississippi	2.47	116.7	288.76
Missouri	3.22	337.92	1088.89
Montana	2.83	54	152.74
Nebraska	3.75	130.05	488.23
Nevada	2.54	127.63	324.01
New Hampshire	2.79	73.36	204.57
New Jersey	3.02	525.17	1587.3
New Mexico	1.75	94.7	165.58
New York	2.65	991.45	2630.51
North Carolina	3.21	563.69	1808.86

2017 NSCH strata and oversample rates, 2015 ACS

State	Oversample rate	BW_S1	BW_S2a
North Dakota	3.52	42.49	149.45
Ohio	3.39	686.03	2327.47
Oklahoma	2.01	188.16	377.4
Oregon	3.53	256.6	905.92
Pennsylvania	3.38	778.07	2630.42
Rhode Island	2.92	52.61	153.5
South Carolina	2.46	253.61	624.64
South Dakota	2.50	54.53	136.36
Tennessee	2.72	374.49	1016.79
Texas	2.46	1450.43	3571.25
Utah	2.42	239.62	579.29
Vermont	2.62	32.35	84.8
Virginia	2.71	579.95	1570.36
Washington	3.55	477.28	1692.78
West Virginia	3.03	81.29	246.5
Wisconsin	3.40	411.98	1400.44
Wyoming	1.54	28.95	44.48

2016 NSCH oversample rates

- Oversample rates are maximized under a constraint that sample variance is only 1.06 times that of a proportional stratified design of a similar cost
- Based on
  - A set of assumptions about response rates from 2015 pretest
  - Estimated child-present rates in each stratum in each state from ACS microdata

2016 NSCH oversample rates

	Oversample rate	$P(C)$	N	N_{S_1}	N_{S_2}
AK	3.8	26.8%	8,468	4,099	4,369
CA	5.3	29.8%	6,407	4,264	2,142
UT	4.4	35.2%	6,126	4,081	2,045

Source: US Census Bureau (2017)

- State-specific oversample rates between 3.4–8.6
- Mean oversample rate between 5–6 depending on weighting

2017 NSCH oversample rates

- Oversample rate was between 3 and 4 depending on measurement

	Oversample rate 2017	Oversample rate 2016	$P(C)$
AK	1.3	3.8	26.8%
CA	2.7	5.3	29.8%
UT	2.4	4.4	35.2%

Source: US Census Bureau (2017)

- Oversample is less severe than in 2016, since the number of negative screeners is expected to be much lower
- Reduces the risk of selection bias between Strata 1 and 2a
- Potentially increases risk of selection bias w.r.t. Stratum 2b under nonsampling

United States Census Bureau