UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

2016 National Survey of Children's Health

Source and Accuracy Statement

U.S. Census Bureau
08/31/2017

1.0 INTRODUCTION

The National Survey of Children's Health (NSCH) is being conducted by the U.S. Census Bureau for the U.S. Department of Health and Human Services' (HHS) Health Resources and Services Administration's (HRSA) Maternal and Child Health Bureau (MCHB). It is designed to provide national and state-level information about the physical and emotional health and well-being of children under the age of 18 living in mailable residential housing units in the United States, their families and their communities, as well as information about the prevalence and impact of children with special health care needs.

This Source and Accuracy Brief/Statement (S&A) provides an overview for the following phases of the 2016 NSCH survey cycle. Hopefully it will enable an understanding of the creation of the data files, as well guidance on their use.

2.0 SAMPLE DESIGN

2.1 Creation of the Sample Frame

The population of interest for the 2016 NSCH is all children under the age of 18, residing in the United States on the date of the survey. Among many other key elements, the survey frame was designed to identify households with children and to provide information about household access to the Internet, which was critical for data collection.

The 2016 NSCH sample frame was developed from two sources: the Edited Master Address File Extract (EDMAFX) created by the Demographic Statistical Methods Division (DSMD) of the Census Bureau; and a file of administrative flags that was created by of the Bureau's Center for Administrative Records Research and Applications (CARRA).

2.1.1 *Use of the Edited Master Address File Extract*

The Census Bureau's Master Address File (MAF) is an accurate and up-to-date inventory of all known living quarters in the United States, Puerto Rico, and associated island areas. It supports

most of the censuses and surveys that the Census Bureau conducts, including the decennial census, the American Community Survey (ACS), and ongoing demographic surveys. The content of the MAF includes mailing and location addresses, unit type attributes, geographic codes for areas such as state, county, census tract, and census block for each living quarters, and source and history data.

The EDMAFX is created at least once every year, specifically for use by DSMD's ongoing demographic surveys. Of importance to the 2016 NSCH is the assignment of a current surveys housing unit validity flag (CSVALDF), resulting from DSMD's filtering rules and processes.

The January 2016 version of the EDMAFX was used in the NSCH sample frame creation…3,142 county-level MAFs rolled up to 51 state-level MAFs, which include the District of Columbia. Only records having CSVALDF=1 (valid housing unit) were kept, with just the unique identification variable MAFID[1] to match to CARRA's file of Administrative Flags.

2.1.2 *Use of CARRA's File of Administrative Flags*

All MAFIDs in the January 2016 MAF-X[2] were appended with flags (e.g., number of children in a household by age group, poverty, and internet access) from data sources such as the Numident and the ACS. This national file was matched to the EDMAFX to produce the sample frame.

2.1.2.1 Processing Overview of CARRA's 2016 NSCH File of Administrative Flags

The frame for all households with children came from three data sources: the Numident; a list of Social Security Number applicants with data updated from various administrative records; and the CARRA kidlink file, a prototype linkage between children and parents based on Census and administrative records.

The Numident is based on all individuals who have been assigned Social Security Numbers. Demographic data from the Numident is updated from federal tax data and various administrative records. There were 75,156,219 children in the December 2015 Numident who would be aged 0–17 years on April 1, 2016.

To identify and sample households containing children in the Numident, the children in the Numident had to be connected to the households in which they live. This was done with the CARRA kidlink file. The CARRA kidlink file uses data from Census survey and federal administrative records to link children Protected Identification Keys (PIKs[3]) to parent PIKs. It identifies the parents of children in the Numident. The source data for the CARRA kidlink file are: the Census Numident, the 2010 Census Unedited File, the IRS 1040 and 1099 files, the Medicare Enrollment Database, the Indian Health Service Database, the Selective Service

[1] Since MAFID cannot be released, similar household ID variables were created and placed on the Screener (HHIDS) and Topical Files (HHID).
[2] CARRA used different extracts of the January 2016 MAF in their processing, specifically the MAF-X and the MAF-ARF.
[3] CARRA uses an anonymous identifier called a PIK to link individuals across datasets while protecting their personally identifiable information.

System, and Public and Indian Housing and Tenant Rental Assistance Certification System data from the Department of Housing and Urban Development. Of these, the IRS 1040 files provided the most significant information.

The MAF Auxiliary Reference File (MAF-ARF) was used to update household location. It links person identifiers to address identifiers using Census survey data and federal administrative data. The source data for the MAF-ARF file are the same as those listed for the CARRA kidlink file.

For each child observation from the Numident, there are four possible MAFIDs: the SSI MAFID, the kid to MAF-ARF MAFID, the child-to-kidlink-to-mother-to-MAF-ARF MAFID, and the child-to-kidlink-to-father-to-MAF-ARF MAFID. Using that order, a single MAFID was allocated. The MAFID match rate was 87.2 percent. The 68,558,710 children associated with a MAFID were then collapsed down to 36,642,194 unique MAFIDS. This implies 1.87 children per household for households assigned a flag.

The MAFID list was then scaled up to the universe of MAFIDs to allow sampling of unflagged households. A merge of the 36,642,194 unique child-flagged MAFIDS with the January 2016 ACS MAF-X file matched 36,609,700 MAFIDS with child flags, removed 32,494 MAFIDS with child flags, and added 159,897,403 MAFIDs without child flags. The resultant file has 196,507,103 valid MAFIDS, of which 36,609,700 MAFIDs include child flags.

See Figure 1 for an overview of the process.

2.1.2.2 Local-area Internet-accessibility

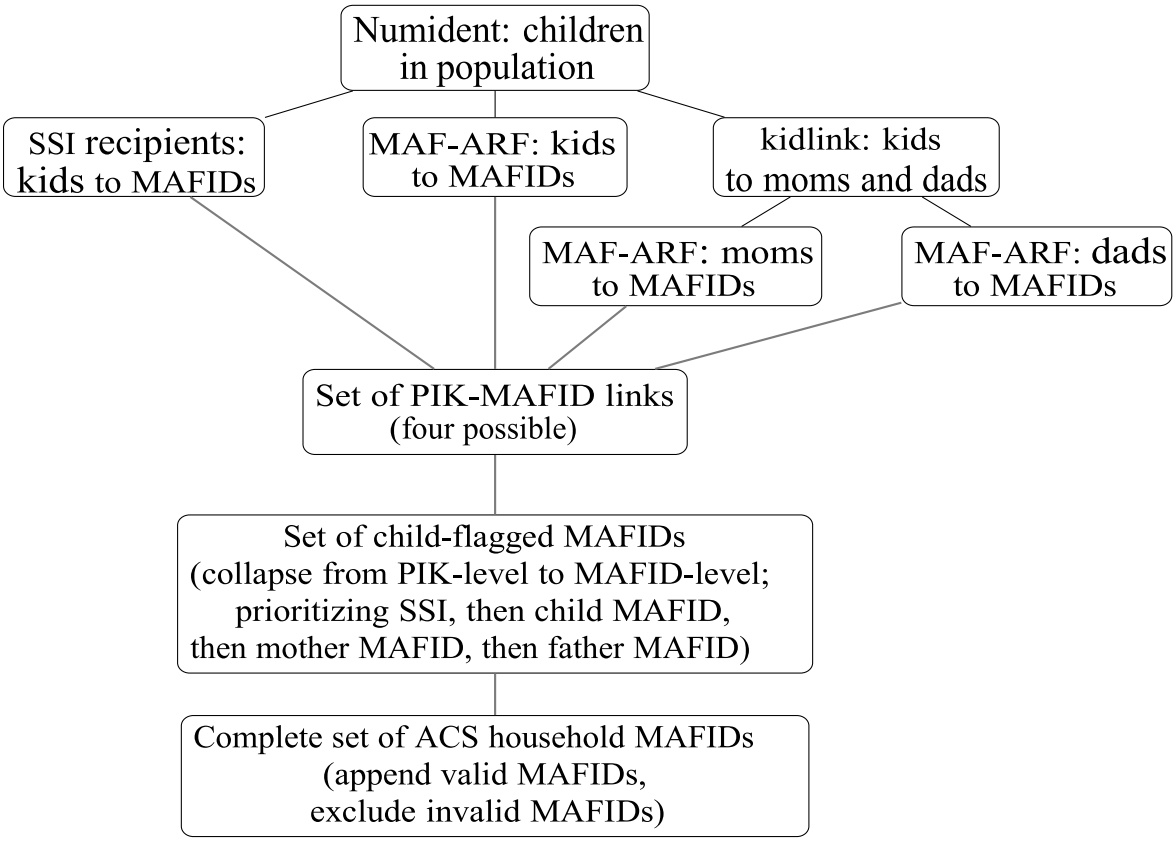
CARRA also created tract-varying internet-accessible household flags. The data came from ACS paradata. Since 2012, ACS respondents have been able to submit survey forms over the Internet. ACS paradata records whether a respondent chose the online option, and it has been summarized at the tract level. The Internet accessibility index is equal to a weighted proportion of the respondents that chose to submit the ACS over the Internet if given the option to do so.

To construct an Internet-access flag, a tract is considered to have low Internet access if the Internet accessibility index is below the first tritile of the tract-level distribution. Ultimately, a variable WEBGROUP was defined as low (L) or high/medium (H).

2.1.2.3 Local-area Household Income Relative to the Poverty Rate

The CARRA file also has a set of poverty variables from the 2014 5-year ACS file. These variables measure the proportion of households with household income in an interval defined by the poverty rate. Ultimately, a variable POVERTY was defined as Y or N from the proportion of households in the block group that have household income less than 150% of the poverty rate (30 percent cut-off).

Figure 1: Illustration of CARRA's File Processing



2.1.3 *Final 2016 NSCH Sample Frame*

The data files detailed in Sections 2.1.1 and 2.1.2 were merged together based on MAFID to create the final sample frame.

2.2 Sampling Strata

Each state had two sampling strata, STRATUM=1 and STRATUM=2, with different state-level sampling intervals for each. The strata were defined by CARRA's presence of children in the household flag. Households flagged as having at least one child under the age of 18 were assigned to stratum 1; all other households were assigned to stratum 2.

2.3 Selection of the Sample Households and Additional Assignments

Recall that the 2016 NSCH sample frame is basically the valid housing units from the MAF, appended with several administrative flags. Table 1 provides the calculated expected sample sizes, by state. Sample sizes were calculated to meet the goal of 1,500 Topical interviews per state, factoring in the expected valid address rate, response rates, and the prevalence of households with children. Addresses in Stratum 1 were sampled at a higher rate than Stratum 2 to increase the number of households with children in the sample while limiting the increase in the variance from the differential sampling rates. The oversampling factor (sampling rate for

Table 1: 2016 NSCH Expected Sample Sizes, by Stratum and by State

State	% Sample w/Children	Total Sample	Stratum 1 Sample	Stratum 2 Sample
Alabama	46.6	7827	4636	3191
Alaska	43.1	8468	4099	4369
Arizona	48.3	7553	4486	3067
Arkansas	45.8	7965	4576	3389
California	57.0	6407	4264	2142
Colorado	55.7	6549	4122	2427
Connecticut	55.9	6526	4361	2164
Delaware	53.0	6890	4552	2338
District of Columbia	43.2	8439	4970	3468
Florida	45.4	8035	4938	3097
Georgia	50.4	7238	4469	2769
Hawaii	38.5	9490	3524	5966
Idaho	53.6	6812	4205	2607
Illinois	55.5	6573	4428	2146
Indiana	55.8	6542	4425	2117
Iowa	61.5	5938	4238	1700
Kansas	57.8	6318	4401	1918
Kentucky	49.4	7380	4410	2969
Louisiana	45.3	8057	4841	3216
Maine	49.4	7385	4484	2901
Maryland	56.0	6511	4363	2148
Massachusetts	58.5	6237	4253	1984
Michigan	58.8	6205	4423	1782
Minnesota	62.2	5871	4217	1654
Mississippi	46.4	7856	4663	3193
Missouri	52.6	6940	4512	2427
Montana	48.0	7605	4352	3253
Nebraska	60.1	6071	4147	1925
Nevada	45.4	8031	4466	3565
New Hampshire	56.5	6460	4243	2217
New Jersey	56.6	6447	4247	2200
New Mexico	42.5	8582	4345	4237
New York	46.1	7918	4307	3612
North Carolina	50.6	7215	4448	2766
North Dakota	51.8	7038	4151	2887
Ohio	57.0	6402	4497	1905
Oklahoma	46.1	7908	4490	3418
Oregon	56.9	6414	4224	2190
Pennsylvania	56.5	6461	4391	2070
Rhode Island	54.4	6713	4359	2354
South Carolina	48.4	7543	4694	2849
South Dakota	49.0	7441	4381	3059
Tennessee	50.4	7241	4500	2741
Texas	51.2	7132	4254	2878
Utah	59.6	6126	4081	2045
Vermont	51.0	7151	4105	3045
Virginia	56.0	6513	4314	2199
Washington	55.4	6583	4269	2314
West Virginia	40.0	9116	4165	4952
Wisconsin	59.5	6136	4267	1869
Wyoming	46.2	7894	4193	3701
National		**364,153**	**222,750 (61.2%)**	**141,402 (38.8%)**

Stratum 1 divided by the sampling rate for Stratum 2) ranged from 3.4 to 8.6 across the states. The total sample size was determined to be 364,150 housing units[4], 222,750 selected from Stratum 1 and 141,400 from Stratum 2. (Note: The expected totals differ from the actual totals as a result of rounding in the sampling process.)

2.3.1 *Process of Selecting Households*

Sampling intervals determined the households selected to be in sample and were calculated for each stratum of each state. The formula is the state-level stratum size on the frame divided by the calculated state-level expected sample size in the stratum.

When determining the random start for each stratum of each state, first the SAS function RANUNI(seed) was used to generate a number from the (0,1) uniform distribution. The returned value was then multiplied by the Sampling Interval to get the random start, or the first record to be in sample for that state and stratum.

2.3.2 *Assignment of Mailing Group, Incentive, and Logo Group to the 364,150 Sample Records*

Screener mailing group (1 or 2), incentive ($0 (control), $2, or $5) and logo group (Census or MCHB-HHS) for each MAFID were assigned randomly across the households that were selected for sample, by state, according to the percentages in Table 2. These assignments for each of the sample records were made before any data was collected.

2.4 Selection of the Sample Children

2.4.1 *Determining Each Child's Eligibility*

A child is an eligible child if their age is less than 18 years.

Table 2: Percent Distribution of Additional Variables

Mailing Group	Incentive	Logo Group
50% → 1	33.3% → $0	50% → Census
		50% → MCHB-HHS
	33.3% → $2	50% → Census
		50% → MCHB-HHS
	33.3% → $5	50% → Census
		50% → MCHB-HHS
50% → 2	33.3% → $0	50% → Census
		50% → MCHB-HHS
	33.3% → $2	50% → Census
		50% → MCHB-HHS
	33.3% → $5	50% → Census
		50% → MCHB-HHS

[4] The total sample size of 364,150 was determined primarily from the available budget.

2.4.2 *Determining the Status of each Eligible Child's Special Health Care Needs*

An eligible child in a household is deemed a child with special health care needs (C_CSHCN=1) if one or more of the following five groups have Screener responses of 'yes' to all of the questions in that group.

If:
Does (fill with CN_NAME) CURRENTLY need or use medicine prescribed by a doctor, other than vitamins? = yes (C_K2Q10=1) **AND**
Is (fill with CN_NAME)'s need for prescription medicine because of ANY medical, behavioral, or other health condition? = yes (C_K2Q11=1) **AND**
Is this a condition that has lasted or is expected to last 12 months or longer? = yes (C_K2Q12=1)

If:
Does (fill with CN_NAME) need or use more medical care, mental health, or educational services than is usual for most children of the same age? = yes (C_K2Q13=1) **AND**
Is (fill with CN_NAME)'s need for medical care, mental health, or educational services because of ANY medical, behavioral, or other health condition? = yes (C_K2Q14=1) **AND**
Is this a condition that has lasted or is expected to last 12 months or longer? = yes (C_K2Q15=1)

If:
Is (fill with CN_NAME) limited or prevented in any way in his or her ability to do the things most children of the same age can do? = yes (C_K2Q16=1) **AND**
Is (fill with CN_NAME)'s limitation in abilities because of ANY medical, behavioral, or other health condition? = yes (C_K2Q17=1) **AND**
Is this a condition that has lasted or is expected to last 12 months or longer? = yes (C_K2Q18=1)

If:
Does (fill with CN_NAME) need or get special therapy, such as physical, occupational, or speech therapy? = yes (C_K2Q19=1) **AND**
Is (fill with CN_NAME)'s need for special therapy because of ANY medical, behavioral, or other health condition? = yes (C_K2Q20=1) **AND**
Is this a condition that has lasted or is expected to last 12 months or longer? = yes (C_K2Q21=1)

If:
Does (fill with CN_NAME) have any kind of emotional, developmental, or behavioral problem for which he or she needs treatment or counseling? = yes (C_K2Q22=1) **AND**
Has his or her emotional, developmental, or behavioral problem lasted or is it expected to last 12 months or longer? = yes (C_K2Q23=1)

2.4.3 *Strategies for Selecting the 2016 NSCH Sample Children (SC_) from the Screener Responses*

For both the Paper and the Web data collection instruments, the sample child was selected from the first four eligible children, after sorting by:

- special health care needs status
  - age (youngest to oldest)
- non-special health care needs status
  - age (youngest to oldest)

In the case of two or three children having the same age and the same special health care needs status, an additional sort by name (A to Z) was implemented. If they also had the same name, e.g., all 'blank', then sorting had no effect.

A special case was children in households that had four or more eligible children. These children were sorted by their special health care needs status, then by name (A to Z), and then sorted by age (youngest to oldest).

A sample child was selected based on the criteria presented in Table 3. The strategies employed allowed for an oversample of both children with SHCNs and children aged 0 through 5 years.

Table 3: Strategies for Selecting the 2016 NSCH Sample Children (SC_)

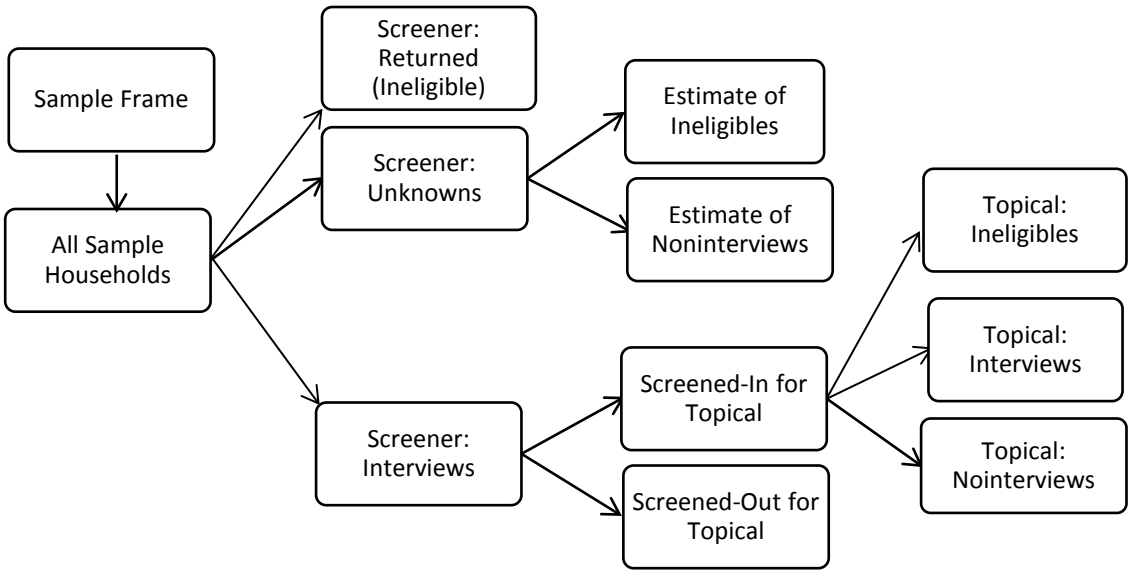
Number of Eligible Children in Household (TOTKIDS_R)	Number of Eligible Non-SHCN (TOTNONSHCN), CSHCN (TOTCSHCN)	% Probability of Selection for Non-SHCN	% Probability of Selection for CSHCN	Notes
1	1,0 or 0,1	100%		Single child is selected.
2	2,0 or 0,2	• If only 1 child is aged 0-5, that child's probability of selection is 62% and the other child's probability of selection is 38%. • Otherwise, each child has an equal chance of selection of 50%.		Includes 60% oversampling of children aged 0-5.
2	1,1	36%	64%	Includes 80% oversampling of CSHCN.
3	3,0 or 0,3	• If only 1 child is aged 0-5, that child's probability of selection is 44% and each of the other two children have an equal chance of selection of 28%. • If 2 children are aged 0-5, each has a probability of selection of 38% and the other child has a probability of selection of 24%. • If all 3 children are aged 0-5 or 6-17, then each child has an equal chance of selection of 33.3%.		Includes 60% oversampling of children aged 0-5.
3	2,1	52% • 26% probability of selection for *each* Non-SHCN child.	48%	Includes 80% oversampling of CSHCN.
3	1,2	22% • 39% probability of selection for *each* CSHCN.	78%	Includes 80% oversampling of CSHCN.
4 or more	Any combination	Before the sort, each of the first 4 children has an equal 25% probability of selection.		Simple random selection of 1 of the first 4 (sorted) children, regardless of Non-CSHCN or CSHCN.

3.0 SURVEY WEIGHTS

3.1 Overview of the Weighting Process

Figure 2 provides a framework for the weighting steps. The weighting process used the data from each phase of the data collection, from both the Paper and Web instruments, to produce final weights for the Screened-in Households, Screener Children, and Interviewed Children.

Figure 2: From Sample Frame to Final Outcome



The weighting process was done by state, with the District of Columbia treated as a state. Weighting for the interviewed children began with the base weight (BW) for each sample household, followed by an adjustment for Screener nonresponse (SNA). Then, the eligible children from the Screener interview cases were raked to population controls (Child-Level Screener Factor=CLSF). A within-household subsampling factor (WHSF) was applied to the Screener interview cases, and an adjustment for Topical nonresponse (TNA) was applied to the Topical interview cases. As a factor for the final weight for interviewed children, a final raking adjustment (RAK) to various demographic controls was performed. The weighting process for all Screener children was a subset of these six factors. Similarly, the screened-in households received a household-level weight, calculated using a small subset of the aforementioned factors as well as a Household Post-Stratification Adjustment (HPSA).

3.1.1 *Baseweight*

The BW for each sample housing unit is the inverse of its probability of selection for the Screener. Each state had two sampling strata with different probabilities of selection for each. If there had been no nonresponse and the survey frame was complete, using this weight would give unbiased estimates for the survey population.

3.1.2 *Screener Nonresponse Adjustment Factor*

The SNA increases the weights of the households responding to the Screener to account for all the households not responding to the Screener.

The count of Screener noninterviews is an estimate of the expected number of eligible households from those cases for which nothing is received. The term eligible here refers to the address belonging to an occupied, residential household. The expected number of eligible cases was estimated by taking the eligibility rate among the known cases and applying it to the unknown cases.

Sixteen Screener weighting cells were defined by the sampling stratum (STRATUM), a block-group poverty measure (yes/no) variable indicating the proportion of households with income less than 150 percent the poverty rate, a measure of internet accessibility (high/medium vs. low), and a Core Based Statistical Area (CBSA) Flag (located within vs. outside of a CBSA).

Within each resultant Screener weighting cell, the SNA was defined as:

$$\left(\frac{\text{weighted sum of Screener interviews + S\_NONINT}}{\text{weighted sum of Screener interviews}}\right)$$

where S_NONINT =

$$\left(\frac{\text{weighted sum of Screener interviews}}{\text{weighted sum of Screener interviews + weighted sum of Screener ineligible households}}\right) \times$$

$$\text{(weighted sum of households with unknown Screener eligibility)}$$

This was the last of the weight processing for Screener households for which there was no Screener interview or interviewed households that indicated no eligible children.

3.1.3 *Household Post-Stratification Adjustment Factor*

All households who indicated on the Screener that there were eligible children present (also called screened-in households) were given a household-level weight. In addition to the BW and SNA, there was an HPSA applied in order to achieve the final screened-in household weight. This factor consisted of ratio adjustments to population controls attained from 2015 ACS data.

Households were put into one of 255 cells depending on their state, race of the selected child, and ethnicity of the selected child if the selected child's race was White. Cells were collapsed as necessary. Within each cell, the HPSA was calculated as the control for the cell divided by the cell's weighted total.

3.1.4 *Child-Level Screener Factor*

All eligible children (at most 4) from the Screener interviewed households were given a Child-Level Screener Weight in order to eventually produce state-level CSHCN prevalence estimates.

This was accomplished through iterative raking to population controls attained from the ACS 2015 single-year estimates.

Raking to the population controls was accomplished using the following three analytical domains of interest, in this order: (Cells were collapsed as necessary.)

- Dimension #1 – State by Child's Race (White, Black, Asian, Other)
- Dimension #2 – State by Child's Ethnicity (Hispanic, non-Hispanic)
- Dimension #3 – State by Child's Gender by Child's Age Group (0-5, 6-11, 12-17)

Each iteration consisted of three ratio adjustments. The ratio adjustments control the weights to the respective dimension control totals. Each ratio adjustment is called a rake. The first rake used the most recent intermediate weight (BW × SNA) as the child's input weight in the raking process. All subsequent rakes used the resulting weight from the previous rake as the input weight. The iterative raking process continued until convergence was met for all cells. Convergence required the following to be true:

(the cell's weighted total – the control for the cell) / control < 10%.

At the end of the process, the CLSF was calculated as the weight after the final iteration divided by the weighted total prior to raking (BW× SNA).

Households where a child was selected from a completed Screener to receive a Topical interview, but become ineligible to complete a Topical were not assigned any further nonzero weighting factors. Examples may include households for which the Screener was received after the final Topical mailing; the child is no longer a resident of the household; etc.

3.1.5 *Within-Household Subsampling Factor*

Weights of the remaining eligible cases were adjusted for the subsampling of children within the households. The value of the adjustment is the inverse of the probability of selection for the selected children. Probabilities varied by the number of children in the household, the presence of children aged 0 through 5, and the presence of CSHCNs. The weights for the selected children now represented all children (at most 4) in the household, and took into account oversampling for CSHCNs and young children. See the details in the previous Table 3.

3.1.6 *Topical Nonresponse Adjustment Factor*

Similar to the SNA, the TNA increased the weights of the households responding to the Topical to account for all of the households not responding to the Topical. These households returned a Screener and went through the subsampling process to select a single child to be the subject of the Topical. If the respondent reached Section H and answered at least 50 percent of key items, then it was considered a Topical interview. A returned Topical that did not meet these conditions was considered a Topical non-interview.

Households were put into one of 96 cells depending on imputed poverty/non-poverty (yes/no), WEBGROUP, number of children (1, 2, 3+), presence of SHCNs of the selected child, and race of the selected child. Within each of the 96 Topical weighting cells, collapsed as necessary:

$$\text{TNA} = \left(\frac{\text{weighted sum of Topical interviews} + \text{weighted sum of Topical Non-interviews}}{\text{weighted sum of Topical interviews}}\right)$$

Households for which there was no Topical interview were not assigned any further nonzero weighting factors.

3.1.7 *Raking Adjustment Factor*

This final step of the weighting process was accomplished through iterative raking to population controls attained from the ACS 2015 1-year estimates and the 2016 NSCH Screener data. Since the process was very similar to that of the CLSF, including checking for convergence, details are omitted in this section. The only significant difference was the dimensions:

- Dimension #1 – State by Household Poverty Ratio (≤1, (1,2], >2)
- Dimension #2 – State by Household Size (2, 3, 4, >4)
- Dimension #3 – State by Respondent's Education <HS, HS, >HS)
- Dimension #4 – State by Selected Child's Race (White, Black, Asian, Other)
- Dimension #5 – State by Selected Child's Ethnicity (Hispanic, Non-Hispanic)
- Dimension #6 – State by Selected Child's SHCN Status (yes/no)
- Dimension #7 – Selected Child's Sex by Single Age, at the National level

At the end of the process, the RAK was calculated as the weight after the final iteration divided by the weighted total prior to raking (BW × SNA × SC_CLSF × WHSF × TNA).

3.1.8 *Trimming Extreme Weights*

The last raking results were checked for extreme values. An extreme value was defined to be one that exceeded the median weight plus six times the interquartile range (IQR) of the weights in each state. These extreme weights were trimmed to this cutoff (six times the IQR of weights in that state). Then, the RAK raking steps were applied again and the new resulting weights were rechecked for extreme values and trimmed as before, continuing as was necessary.

Table 4 shows the distribution of the weights, by state, after the fourth trimming and re-raking procedure and before the last and final trimming step.

Table 4: Summary of Last Raking Result before Final Trimming

STATE	MIN	Q1	MEDIAN	Q3	MAX	IQR	CUTOFF median+6*IQR	No. of Extremes
Alabama	196.2	556.5	872.7	1613.4	7268.3	1056.9	7214.2	6
Alaska	29.5	93.0	155.2	267.7	1208.1	174.7	1203.7	10
Arizona	211.7	646.9	1124.2	2170.0	10957.2	1523.1	10262.7	13
Arkansas	100.7	321.0	574.5	1005.3	4705.3	684.3	4680.3	5
California	1183.4	3430.6	5613.5	10834.9	51626.4	7404.3	50039.1	20
Colorado	219.5	518.2	771.0	1288.3	5693.0	770.1	5391.9	18
Connecticut	101.0	289.9	476.5	834.9	3789.8	545.0	3746.4	7
Delaware	30.1	84.7	146.9	268.6	1265.7	183.9	1250.3	4
District of Columbia	8.7	29.4	56.1	131.2	692.9	101.8	666.9	13
Florida	637.0	1704.0	2822.4	5340.5	24866.2	3636.5	24641.4	6
Georgia	382.4	1241.6	2097.1	3742.8	17130.3	2501.2	17104.3	2
Hawaii	45.9	124.9	200.1	323.1	1395.5	198.2	1389.3	5
Idaho	19.9	176.5	291.3	505.8	2277.6	329.2	2266.7	4
Illinois	445.1	1130.0	1725.2	3137.8	14022.1	2007.7	13771.5	12
Indiana	289.9	792.1	1224.4	1983.8	8437.8	1191.8	8375.1	7
Iowa	139.0	296.7	459.8	761.9	3304.2	465.3	3251.4	9
Kansas	134.8	317.1	473.8	878.3	3894.7	561.2	3841.3	8
Kentucky	97.1	482.1	758.2	1346.2	5993.2	864.1	5942.9	6
Louisiana	159.9	642.7	1125.0	1957.4	9209.3	1314.7	9013.0	6
Maine	64.3	123.2	182.1	288.1	1180.8	164.8	1171.1	12
Maryland	210.7	525.6	850.6	1428.2	6435.8	902.6	6266.2	10
Massachusetts	207.5	470.0	800.2	1347.5	6399.4	877.5	6065.4	13
Michigan	405.4	853.1	1315.5	2210.2	9687.9	1357.2	9458.5	14
Minnesota	196.5	383.5	590.0	1191.1	5490.8	807.6	5435.5	10
Mississippi	184.7	434.1	787.4	1410.1	6780.2	976.0	6643.3	7
Missouri	304.9	622.3	949.6	1591.7	6756.9	969.3	6765.6	0
Montana	40.6	92.1	142.2	263.4	1177.3	171.4	1170.3	8
Nebraska	86.2	193.2	335.8	528.2	2351.4	334.9	2345.3	1
Nevada	111.7	332.3	557.4	1028.0	4783.0	695.7	4731.8	10
New Hampshire	52.9	125.2	193.9	302.0	1260.1	176.8	1254.5	12
New Jersey	340.2	753.0	1134.6	2201.2	10021.9	1448.1	9823.4	16
New Mexico	81.9	247.2	398.6	721.9	3301.4	474.7	3246.8	11
New York	787.1	1828.4	2924.9	5098.7	23112.5	3270.3	22546.5	10
North Carolina	267.4	934.8	1476.0	2629.1	12940.1	1694.2	11641.4	24
North Dakota	32.8	73.7	118.0	212.2	948.9	138.5	948.8	1
Ohio	360.2	995.3	1710.4	3063.6	14169.3	2068.4	14120.6	4
Oklahoma	183.7	566.0	930.1	1482.3	6479.4	916.3	6427.8	8
Oregon	132.3	292.0	484.0	921.1	4316.6	629.1	4258.8	12
Pennsylvania	383.6	1019.2	1616.3	2764.3	12333.1	1745.1	12087.0	16

STATE	MIN	Q1	MEDIAN	Q3	MAX	IQR	CUTOFF median+6*IQR	No. of Extremes
Rhode Island	38.3	95.8	155.5	259.2	1214.0	163.4	1135.8	21
South Carolina	150.3	451.9	841.3	1462.1	6948.6	1010.3	6902.8	2
South Dakota	30.4	79.6	128.7	210.2	932.0	130.6	912.2	21
Tennessee	236.2	703.0	1091.6	1929.0	8624.5	1226.0	8447.6	16
Texas	893.1	3300.9	5557.8	10273.3	48614.0	6972.4	47392.2	7
Utah	161.0	388.8	642.5	986.2	4230.2	597.4	4226.8	2
Vermont	15.5	47.4	70.5	117.6	493.0	70.2	491.9	7
Virginia	245.9	722.3	1095.4	1821.9	7776.8	1099.6	7692.8	17
Washington	258.3	548.3	847.4	1746.1	8314.3	1197.7	8033.8	17
West Virginia	46.2	219.8	346.0	593.5	2597.7	373.6	2587.8	5
Wisconsin	220.3	457.1	679.4	1134.7	4846.2	677.6	4745.1	32
Wyoming	31.6	64.8	119.8	192.9	914.4	128.1	888.4	15

As shown by the low number of extremes in the final column and the proximity of the maximums to the cutoffs, by state, it was decided to perform the final trimming at this point.

3.2 Final Weights Produced

Selected Child Weight (Topical) = FWC = BW × SNA × SC_CLSF × WHSF × TNA × RAK
Child Weight (Screener) = C_FWS = BW × SNA × C_CLSF
Household Weight (Screener) = FWH = BW × SNA × HPSA

3.3 Population Controls

The ACS is an ongoing national survey that samples approximately 3.5 million addresses annually, averaging about 290,000 addresses per month. These data are collected continuously throughout the year to produce annual population and housing estimates. The survey covers the resident population of the United States and Puerto Rico for people living in housing units and group quarters. (Note that the 2016 NSCH weighting cells only used the resident population of the United States for people living in housing units.)

The survey produces critical information for small areas and small population groups – it is the only source of information for many of its topics in these small areas.

Two different sets of estimates, with weights, are released each Fall in the form of single-year (12 months of data) and 5-year (60 months of data) datasets. The 2016 NSCH weighting cells used the 2015 single-year ACS population controls.

3.4 Checks to Guide the Use of the Three Final Weights

Using the assortment of Population Controls in the Attachment[5] and the final weights in the NSCH files, the following are a few checks that the data user can do to more fully understand the use of the final weights:

- Check that the sum of the household weights for Screener interviews matches the control for each state (Column 2).
- Check that the sum of the Screener weights for children match the controls for each state (Column 3).
- Check that the sum of the Screener weights for female children match the controls for each state (Column 4).
- Check that the sum of the Topical weights for children with Poverty Ratio <=1 match the controls for each state (Column 5).
- Check that the sum of the Topical weights for children with respondent education <HS match the controls for each state (Column 6).
- Check that the sum of the Topical weights for children with SHCNs match the controls for each state (Column 7).

4.0 CALCULATING SAMPLING ERROR OF SURVEY ESTIMATES

4.1 Description of Sampling Error

The NSCH estimates are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire and instructions. This difference is known as sampling error and can be estimated from the survey data. While the simplest calculations of sampling error assume simple random sampling, these will underestimate the sampling error for the 2016 NSCH. This is because different sampling rates were used across the two sampling strata, as well as across states, which moved the NSCH design from a simple random sample to a complex sample design.

Standard errors indicate the magnitude of the sampling error and can be used to construct confidence intervals around the survey estimates. By calculating the confidence intervals for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

4.2 Estimating Sampling Error for the 2016 NSCH

Standard errors for the NSCH estimates can be obtained using the Taylor Series approximation method, which is available in software packages such as SAS, Stata, and SUDAAN. The sampling strata are identified by state and the child stratum flag, and the Primary Sampling Unit (PSU) is the household.

[5] The attachment was previously presented both as an Excel spreadsheet and SAS dataset.

For SAS, the following statements are used:

- proc surveyfreq (or proc surveymeans or proc surveyreg)
- strata FIPSST and STRATUM
- cluster HHIDS (for the Screener) HHID (for the Topical)
- weight FWH (household weight), C_FWS (child weight), FWC (selected child weight)

For Stata the following statements are used:

- svyset strata FIPSST and STRATUM
- svyset psu HHIDS (for the Screener) or HHID (for the Topical)
- svyset pweight FWH (household weight), C_FWS (child weight), FWC (selected child weight)

For Stata, the two stratum variables need to be combined into a single variable.

For SUDAAN the following statements are used:

- proc …. design = WR;
- nest FIPSST STRATUM (HHIDS for the Screener or HHID for the Topical) / psulevel=3
- weight FWH (household weight), C_FWS (child weight), FWC (selected child weight)

For SUDAAN, the data file needs to be sorted by FIPSST and STRATUM, and then HHIDS (for the Screener) or HHID (for the Topical). HHID, HHIDS, FIPSST and STRATUM must be converted from character to numeric variable type.

5. Supporting Material

U.S. Census Bureau. Center for Administrative Records Research and Applications. "A sample frame built from administrative records for the National Survey of Children's Health by Keith Finlay." Unofficial document from Keith Finlay, dated April 1, 2016.

U.S. Census Bureau. "Sampling Specifications for the 2016 National Survey of Children's Health, including Creation of the Sample Frame." Forthcoming finalized memorandum from James B. Treat to Barry F. Sessamen.

U.S. Census Bureau. "Subsampling Specifications for the 2016 National Survey of Children's Health." Forthcoming finalized memorandum from James B. Treat to Barry F. Sessamen.

U.S. Census Bureau. "Weighting Specifications for the 2016 National Survey of Children's Health." Forthcoming finalized memorandum from James B. Treat to Barry F. Sessamen.

Attachment: Various Population Controls, by State

State	# of HHs with Children	# of Children	# of Female Children	# of Children in a HH with Poverty Ratio < 1	# of Children in a HH with Respondent's Education < HS	# of CSHCNs
Alabama	565085	1106293	541863	304160	149156	235640
Alaska	86148	185534	90003	30144	14758	29871
Arizona	760908	1618861	795117	417135	278248	317297
Arkansas	358773	702195	339875	200007	94456	162207
California	4511468	9097914	4457164	2017538	1938500	1410177
Colorado	657324	1253909	614607	196581	145300	221942
Connecticut	399791	761030	371405	118442	63260	155250
Delaware	105226	203545	99834	41241	24494	46612
DC	63894	116927	57216	30598	16471	21398
Florida	2033763	4088990	2010949	986389	489055	891400
Georgia	1254477	2497644	1221466	633645	348110	494534
Hawaii	140859	309711	152346	47136	22438	42121
Idaho	192805	430251	210144	81442	42808	75294
Illinois	1495835	2954616	1449260	595338	342172	573196
Indiana	786632	1575789	774256	352321	185902	356128
Iowa	366840	726089	356118	116696	54519	128518
Kansas	352514	718852	352525	129936	78541	147365
Kentucky	527232	1008114	493168	275949	116382	256061
Louisiana	544108	1114329	540716	324897	143304	264096
Maine	139007	255251	123323	49703	10513	58708
Maryland	718007	1345020	657847	194122	131211	250174
Massachusetts	750880	1381629	675866	217341	113602	287379
Michigan	1120119	2201820	1077024	522406	211917	444768
Minnesota	654414	1278646	622641	181141	99757	228878
Mississippi	369092	725662	357677	232456	89874	177062
Missouri	707372	1386641	677848	299471	129371	316154
Montana	109289	224671	110469	46617	14935	41789
Nebraska	232670	468019	229296	82967	43895	78159
Nevada	317492	667452	328166	146966	122355	102120
New Jersey	1059720	1992686	975257	327516	196297	350713
New Mexico	226731	498200	243524	145745	86698	98145
New York	2158715	4187616	2042858	961263	611529	766334
North Carolina	1190059	2281906	1114186	557252	304983	492892
North Dakota	88360	170673	84025	22463	8501	30892
Ohio	1363463	2625727	1283236	589334	252993	598666
Oklahoma	475848	957560	465100	224446	122573	210663
Oregon	440131	858234	420722	186181	106003	158773

State	# of HHs with Children	# of Children	# of Female Children	# of Children in a HH with Poverty Ratio < 1	# of Children in a HH with Respondent's Education < HS	# of CSHCNs
Pennsylvania	1399673	2680763	1310799	547881	262371	517387
Rhode Island	116904	211090	103094	42391	24696	45595
South Carolina	550782	1086138	527105	275179	127621	219400
South Dakota	102643	208561	100840	40462	17125	32744
Tennessee	778154	1490040	732256	375335	173876	283108
Texas	3485107	7194057	3529309	1707820	1457510	1309318
Utah	382658	908856	442091	124137	70621	149052
Vermont	64567	119460	58200	18216	6936	25206
Virginia	986854	1865835	912334	296449	165695	391825
Washington	837343	1609659	787132	271754	179417	299397
West Virginia	201754	378213	187295	100175	34805	91149
Wisconsin	671970	1289359	628341	228768	109730	246268
Wyoming	69007	138841	68519	19752	11732	28185
National	37113984	73421397	35932722	15967230	9858517	14212251