2015-2017 National Survey of Family Growth (NSFG): Weighting Design Documentation

This document is a detailed supplement to another document that serves as a brief, summary description of all aspects of the methodology and operations for the 2015-2017 data release. The summary document is referred to as "summary methodology document" below and is entitled "2015-2017 National Survey of Family Growth (NSFG): Summary of Design and Data Collection Methods."

1. Executive Summary

The target population for the National Survey of Family Growth (NSFG) consists of all non-institutionalized women and men ages 15-49, living in households, and whose usual place of residence is the 50 United States or the District of Columbia. This document summarizes the steps taken in developing weights for the 2-year data files for 2015-2017, which represent the third data file release from the NSFG Continuous 2011-2019 fieldwork period. This two-year period covers the 17th through 24th quarters of the overall planned eight years of data collection. See "2011-2013 National Survey of Family Growth (NSFG): Weighting Design Documentation" and "2013-

2015 National Survey of Family Growth (NSFG): Weighting Design Documentation" for analogous reports for the first eight quarters (September 2011 through September 2013) and the second eight quarters (September 2013 through September 2015) comprising the first and second data releases from the 2011-2019 period of continuous interviewing. Most of the details of the process are the same for the 2015-2017 period as those used in the 2011-2013 and 2013-2015 periods. One significant change was made for the 2015-2017 sample design. Beginning in Sept 2015, or Quarter 17 of fieldwork, the age eligibility requirements were expanded from 15-44 years to 15-49 years.

This report documents the following components of the weights:

- Base probability of selection weights
- Nonresponse adjustments
- Post-stratification adjustments

The base probability of selection weight is based on the five stages of selection in the sample design. The nonresponse adjustment is done for two stages: nonresponse to a screening interview designed to determine eligibility and nonresponse to the request for eligible persons identified in the screening stage to complete a main interview. Post-stratification adjusts the set of respondents based upon known totals for subgroups of the population. A further step in the development of the weights is to trim extreme weights. Extreme weights have been trimmed at several of the steps of the weight development process by smoothing weights across categories of respondents with similar characteristics. This latter type of smoothing is described in each section of the weight construction process.

2. Probabilities of Selection

2.1 The Role of Selection Weights

A base or starting strategy with most survey sample designs is to consider a representative sample, one that is a "scale model" of the population from which the sample is to be selected. However, smaller groups in the population may have too few cases in the sample to provide adequate precision for those groups. Survey sample designs such as the NSFG thus deliberately over-represent smaller groups in the sample. This over-representation is accomplished through the use of varying probabilities of selection. Over-represented groups have higher sampling rates than other groups.

For example, non-Hispanic black women represented approximately 7.5 percent of the U.S. household population 15-49 years of age in 2017. Yet, for purposes of improved precision for non-Hispanic black women, NSFG 2015-2017 chose the sample in such a way that these women account for about 13.4 percent of all respondents in the sample. Similar kinds of over-representation have occurred for non-Hispanic black men, Hispanic women and men, and teenagers of all races. By extension, the over-representation of these groups means than non-Hispanic men and women ages 20-49 are under-represented in the samples. As in previous NSFG data files, "sampling weights" are needed to adjust for these different rates and this over-representation. Without appropriate weighting, resulting estimates from the survey could be subject to substantial bias.

In addition to these over-sampling rates, other factors within the design affected the sampling weights. For a full description of the NSFG design, please see the Sample Design documentation.

The final NSFG sample weights are comprised of four major components: an adjustment for unequal probability of selection, a unit nonresponse propensity adjustment, a post-stratification factor, and a weight trimming step. The adjustment for unequal probability of selection is discussed in the next section. The procedures to develop the latter three components of the final sampling weights are described in the following sections.

For purposes of description, it may be useful to observe that the final weight can be interpreted as the number of persons in the population that an individual NSFG respondent represents. A final weight for a teenage Hispanic female of 2,000 means that this sample respondent represents herself and 1,999 other similar women in the population. The NSFG 2015-2017 final weights are values greater than 1, and when summed across a subgroup or the total sample are expected to provide an estimate of the total number of persons in that subgroup in the U.S. household population.

2.2 Overview of Stages of Selection

The sample was selected in five stages. In the **first stage**, Primary Sampling Units (PSUs) were selected. PSUs are Metropolitan Statistical Areas (MSAs), counties or groups of counties. The 50 United States plus the District of Columbia were divided into 2,149 PSUs on the sampling frame. Of these, 366 are MSAs and 1,783 are non-MSA

PSUs that include one or more counties. The PSUs were stratified according to attributes such as Census Division, MSA status, and size. One or two PSUs were selected with Probability Proportionate to Size (PPS) from each stratum. These PSUs were selected using systematic sampling when more than one PSU was selected. The PPS selection method assigns higher probabilities to PSUs with larger populations. The first stage selection probabilities are inversely related to the probabilities of selection at the second and third stages of selection such that sampling rates are approximately equal for all households within a sampling domain (designed for oversampling black and Hispanic persons, see Table 1). In general, large PSUs have lower within-PSU sampling rates while smaller PSUs have higher within-PSU sampling rates such that households in the same domain but different PSUs have approximately the same chance of being selected. The largest PSUs were selected with probability equal to 1.0 since any national sample of this size should include them. These PSUs are known as "certainty selections" or "self-representing" PSUs. These self-representing PSUs are in strata with only one PSU per stratum. For the 2015-2017 dataset, there were nine such self-representing strata. There were also eight PSUs from strata that would be self-representing after three years, but are not self-representing in a two-year dataset. The remaining PSUs were from an additional 48 non-self-representing strata.

Table 1. NSFG Sample Domain Definitions and Characteristics

Domain	Definition	Total Households	Est. Proportion Black	Est. Proportion Hispanic	2011-2019 Rate/Domain 1 Rate
1	<10% HH Black, <10% HH Hispanic	65,009,685	0.018	0.022	1.0
2	>=10% HH Black, <10% HH Hispanic	19,871,976	0.426	0.029	2.6
3	<10% HH Black, >=10% HH Hispanic	20,270,438	0.026	0.380	2.3
4	>=10% HH Black, >=10% HH Hispanic	11,564,193	0.301	0.299	2.5

In order to facilitate the oversample of subgroups defined by race and ethnicity, the measures of size for the PSUs were a weighted combination of household counts. All Census Block Groups were classified into four sampling "domains." Households in domains 2, 3, and 4 were given a higher weight so that they would have a higher chance of being selected than those in domain 1. These weighted measures of size were used in both the first and second stages of selection.

The **second stage** of selection was to select neighborhoods within PSUs. These selections are called Secondary Sampling Units (SSUs or segments) and are composed of one or more Census blocks with a minimum measure of size equal to 50. The minimum size requirement insures that within-SSU samples are large enough to support efficient interviewer travel. SSUs are selected with PPS. The samples were selected using systematic sampling. The measures of size for these PPS selections are weighted measures of size such that SSUs with larger non-Hispanic black and Hispanic populations received higher probabilities of selection.

SSUs in domains 2, 3, and 4 have relatively higher combined PSU, SSU, and housing unit selection rates. These weighted measures of size and sampling rates are set such that interviews with black and Hispanic respondents each constitute 20% of all interviews. The relative rates of oversampling each domain are given in the last column of Table 1.

Each PSU was assigned one or two interviewers based on its relative size. For each interviewer, 12 SSUs were selected each year. These SSUs were then randomly divided into four groups. One group was released each calendar quarter.

In preparation for the **third stage** of selection, ISR interviewers updated commercially-available lists of housing units for SSUs where these lists were available or, alternatively, created such a list from scratch where they were not available. Once these lists were updated, a sample of housing units was selected.
The selected units were contacted by ISR interviewers to conduct a brief household screener to determine if any members of the household were eligible for the NSFG interview. In households with eligible persons, the **fourth stage** of selection involved selecting one of the eligible persons. In households with only one eligible person, that person was selected. In households with more than one eligible person, one of these persons was randomly selected for inclusion in the survey. The within-household selection rates were set up such that 18.2% of all interviews would be with teens aged 15-19 and 55% of all interviews would be with females. The requirement

for oversampling teens creates relatively extreme weights for those adults who live with teens and were selected for the interview.

As was done in NSFG surveys for 2006-2010, 2011-2013, and 2013-2015, the 2015-2017 NSFG also used a two-phase sampling approach within each data collection quarter as a **fifth stage** of selection. During week 10 of each quarter, a subsample of active cases was selected for continued follow-up. In weeks 11 and 12, this subsample received a special mailed incentive and the interviewers focused effort on the fewer cases left in the subsample. Details of this two-phase design are discussed in more detail later in this report, and described in Lepkowski et al. (2013). Also see the section on "Use of Incentives" in the summary methodology document.

Due to the continuous design, the 2015-2017 NSFG also had to account for the distribution of the **sample across time** in addition to the distribution of the sample across geography. The sample was designed so as not to confound characteristics of the sample with time. For example, if PSUs were rotated over time using Census Division (i.e., release PSUs from Division 1 in Year 1, Division 2 in Year 2, and so on), then division and time would be completely confounded. If that were the PSU release schedule, then it would not be possible to determine whether changes in estimates were due to differences across Census Divisions or temporal changes in the population. For this reason, PSUs were randomly allocated across years in such a manner that each year's sample is a nationally representative sample. Given the numbers of interviews conducted per year, a single year would not produce precise estimates, but when two or more years of data are combined, more precise estimates are possible.

3. Primary Stage Unit (PSU) Probabilities

This section describes the selection probabilities of primary stage units (PSUs). It begins with a description of the development of "weighted measures of size," and then how the PSU probabilities of selection were calculated. This section further discusses how the allocation of PSUs across time modifies these probabilities of selection using the concept of "probability of being released."

3.1 Weighted Measure of Size and Selection Probability

The PSUs of this multi-stage area probability sample were selected with Probabilities Proportionate to Size. A weighted Measure of Size (MOS) is a measure whereby subpopulations for which an oversample is desired were multiplied by a weighting factor that increases the probability of selection for units in that domain. This allowed us to oversample particular subgroups in the population. A weighted measure of size $M_{h\alpha\beta}$ for the β^{th} area segment in the α^{th} PSU in stratum *h* was created as follows. If a block is in a block group with at least some threshold proportion of the population being black or Hispanic, then the count of occupied housing units in that block was multiplied by a factor set such that targeted oversamples for blacks and Hispanics was achieved, based on the four domains defined above (see Table 1). For all other blocks, the measure of size $M_{h\alpha\beta}$ is the 2010 Census occupied housing unit count for the block.

Having determined these block-level composite measures of size, the next step was to sum them to the PSU level across all blocks in the PSU to obtain a PSU level measure of size $M_{h\alpha}$, and the PSU measures of size summed to a stratum size M_h. Within a PSU stratum, a single PSU was selected with probability proportionate to the composite measure of size in most cases, or $M_{h\alpha}/M_h$. In self-representing strata, where the PSU is so large that it will come into the sample with certainty, the probability of selection is 1.0. In all other strata, it is less than 1.0. With single selections per stratum, the following equation can be used to calculate the PSU selection probability. The notation below is simplified by suppressing the stratum and PSU indices from the left-hand side of equation 1:

$$\pi_1 = \frac{M_{h\alpha}}{M_h} \qquad (1)$$

3.2 Probability of Being Released by Year

A sample of 213 PSUs was selected for NSFG 2011-2019, with the expectation of release across this survey period. These 213 PSUs included 21 self-representing PSUs and 192 nonself-representing PSUs. In addition, 2

PSUs were selected to represent Alaska and Hawaii. The release of the PSUs across time has been controlled such that the number of years in the sample is roughly proportional to the size of the PSU. Three PSUs were large enough to be included every year (SR PSUs 1, 2, and 3). There were six additional PSUs (SR PSUs 4-9) that were large enough to be included in two out of three years. These were in groups of two PSUs in a single "super-stratum," or groupings of similar strata. For example, SR PSUs 4 and 5 were in a single "super-stratum." An additional twelve PSUs (SR PSUs 10-21) were included as self-representing, for inclusion in one year every three years. These PSUs were in groups of three PSUs that formed a single "super-stratum." For example, SR PSUs 10, 11, and 12 were a "super-stratum" with one of these PSUs released every third year. That is a total of 21 self-representing PSUs. Each year, an additional 24 nonself-representing PSUs were to be released. During 2015-2017, nine PSUs were self-representing; eight were self-representing after three years (but were not self-representing after only two years), and 48 were nonself-representing. This was a total of 65 PSUs.

A general notation for the probability of being released is based on the number of PSUs selected, g_l, within "super-stratum" *l* and the number of PSUs released $g_{l,yr5-6}$ in years one and two in super-stratum *l*. The probability of being released for a PSU, is therefore, the following:

$$\pi_{1,yr5-6} = \frac{g_{l,yr5-6}}{g_l} \qquad (2)$$

For the top nine self-representing PSUs, the probability of being released in any two-year interval is 1.0. For self-representing PSUs 10-21, the probability of being released in any two-year interval is 2/3, since eight of the twelve PSUs were systematically selected for that time interval. For the nonself-representing PSUs, the probability of being selected for any two-year interval is 48/192=0.25. Therefore the probability that a PSU was released during the two-year interval 2015-2017 can be expressed with the following equations.

SR PSU 1-9 $\qquad \pi_{1,yr5-6} = 1.0$

SR PSU 10-21 $\qquad \pi_{1,yr5-6} = \frac{8}{12} = 0.667$

NSR PSUs $\qquad \pi_{1,yr5-6} = \frac{48}{192} = \frac{1}{4} = 0.25$

The probability of a PSU being released in years 5 or 6 is the following:

$$\pi_{1,yr5-6} = \frac{g_{l,yr5-6}}{g_l},$$

where $g_{l,yr5-6}$ is the number of PSUs released in super-stratum *l* in years 5 or 6. The probabilities for SR PSUs 1-9, SR PSUs 10-21 and NSR PSU is the same as given above.

4 Secondary Stage Unit (SSU) Probabilities

For the 2015-2017 NSFG, the choice of 12 second-stage units that could be allocated in sets of three across quarters in a calendar year was retained because it yields a good balance between cost-efficiency and sampling variance. In some cases, a larger PSU required larger sampling rates. When this could not be accommodated by releasing the PSU in multiple years, this was accomplished by adding another interviewer and selecting 24 second-stage units.

4.1 Second-Stage Selection

The second-stage units (SSUs), termed "area segments," are Census Blocks or combinations of Census Blocks. Within each sample PSU, segments were implicitly stratified by ordering the list of segments by the density of black and Hispanic households (for example, from high to low, within block groups) and systematically selected with probabilities proportionate to weighted measures of size. The construction of these weighted measures of size was described in the previous section.

A measure of size was then calculated for each segment. A domain-specific multiplier (see Table 1) was used to assign higher probabilities of selection to segments in high-density minority domains (i.e., domains 2-4). The result of these weighted measures of size is a disproportionate allocation of the area segment selections to high minority domains. This approach yields sampling rates for high density segments that are 2.3 to 2.6

times larger than those for other segments. The following equation shows the selection probability for the SSU:

$$\pi_2 = \frac{d_{h\alpha} M_{h\alpha\beta}}{M_{h\alpha}} \qquad (3),$$

where $d_{h\alpha}$ is the number of segments selected in the α^{th} PSU in stratum h . Usually 12 segments were selected, sometimes more. For example, in SR PSU 1, there were two interviewers and 24 area segments for each of the eight years of data collection.

In each calendar quarter within a PSU, one-quarter of the segments allocated to each PSU in the yearly sample were selected to be released. Over an entire year, approximately 456 segments were released in the sample. The sample for NSFG 2015-2017 included 65 PSUs and 912 segments (456 per year x 2 years = 912. During two quarters in 2016 (quarter 19 and to a lesser extent quarter 20), sample sizes were reduced relative to the expectation in order to reduce costs. This meant that 9 segments that were sampled for this time period were not released. These 9 were a random subset of sampled segments in several PSUs.) In PSUs that appear only in one year, the probability of an SSU being released in a two-year dataset is 1.0. However, for PSUs that appear in multiple years, the probability of being released needs to be calculated. Recall that $d_{h\alpha}$ denotes the number of segments selected in the α^{th} PSU in stratum *h* for the entire eight-year sample. If the number of segments to be released in years one and two is denoted $d_{h\alpha,yr5-6}$, then the probability of release can be calculated using the following equation:

$$\pi_{2,yr5-6} = \frac{d_{h\alpha.yr5-6}}{d_{h\alpha}}$$

If all the segments are released in years 3 and 4, then the probability of being released is 1.0.

Across the two stages of selection, the probability of selection is $\pi_1 \times \pi_2 = \frac{M_{h\alpha}}{M_h} \times \frac{d_{h\alpha} M_{h\alpha\beta}}{M_{h\alpha}}$. With the composite measures of size, relatively more high density segments are selected for housing unit sampling and screening. These probabilities of being selected were further modified by the probabilities of being released to form a combined probability of being selected and released in years 5 and 6 (2015-2017):

$$\pi_1 \times \pi_{1,yr5-6} \times \pi_2 \times \pi_{2,yr5-6} = \frac{M_{h\alpha}}{M_h} \times \frac{g_{l,yr5-6}}{g_l} \times \frac{d_{h\alpha} M_{h\alpha\beta}}{M_{h\alpha}} \times \frac{d_{h\alpha.yr5-6}}{d_{h\alpha}}.$$

5 Tertiary Stage: Housing Unit Probabilities

The NSFG began from an overall design that was targeted to achieve a minimum of 5,000 interviews per year, with oversamples of females, teens, blacks, and Hispanics. In addition, the labor model for interviewers on this survey requires that they are provided sufficient work for each interviewer to work, on average, 30 hours per week. Under this approach, the sample sizes for each interviewer were allowed to vary such that they had, in expectation, a large enough sample of housing units to sustain 360 hours of work each quarter. Beginning in Quarter 13, a sample design change was implemented that was meant to increase the percentage of screened households that contain an eligible person. This was accomplished by stratifying housing units based on a prediction of whether the unit contained an eligible person. The model was selected and estimated using data from previous quarters where the binary eligibility outcome was measured. Key predictors in this model included commercial data that estimated whether an eligible person was in the household. The predicted probability of there being an eligible person in the household was used to create strata and then oversample the stratum or strata with higher expected eligibility. The method for setting the housing unit selection rates is described below.

5.1 Third-Stage Selection of Housing Units

The third stage random selection of housing units was made from the segment housing unit list. As a first step in setting the sampling rates for this stage, a single within-segment sampling rate was set. This initial rate is a function of the efficiency of the interviewer, but did not differentiate rates across housing units within the same segment based upon predicted eligibility. More efficient interviewers would have relatively higher sample sizes to insure that every interviewer had enough sampled housing units that they could work 30 hours every week for 12 weeks.

The beginning sampling rate was set to be equal probability selection method (EPSEM) within domain. This rate can be calculated using the following formula:

$$\pi_3 = \frac{\pi_d}{\pi_1 \times \pi_2}$$

Here, π_d is the overall sampling rate for the domain and π_1 and π_2 are the PSU and SSU selection probabilities (described in the previous section). The values for π_d are given in Table 2.

Table 2. Domain-level Sampling Rates

Domain	**Overall Domain Sampling Rate (π_d)**
1	0.000465968
2	0.001211516
3	0.001071726
4	0.001164919

Once these rates (π_3) had been set and the listing of housing units completed, a preliminary expected sample size was calculated. This sample size is the number of housing units listed ($HU_{\alpha\beta t}$) multiplied by the initial rate (π_3) and is denoted ($L_{\alpha\beta t} = \pi_3 \times HU_{\alpha\beta t}$). This preliminary sample size was modified by a multiplier designed to produce a sufficient sample size for a given interviewer efficiency.

The sufficient sample size for an interviewer was calculated at the PSU level. Within an expected 360 hours in a 12-week period, interviewers updated or prepared "scratch" listings for the segments allocated in the next calendar quarter, screened selected lines, and conducted main interviews. Interviewers had varying survey conditions in their work assignments that made them more or less efficient within the 360 hours available. The conditions varied by the nature of the communities in which they worked, which in turn affected parameters such as the number of hours required to complete an interview (i.e., the hours per interview, or for the α^{th} PSU at calendar quarter *t*, HPI_{at}); the housing unit occupancy rate (O_{at}); the proportion of occupied housing units with one or more persons ages 15-49 (the eligibility rate, E_{at}); the proportion of the sample that is either completed during phase one or will be retained for phase 2 (the subsampling rate $\hat{S}_{at}$); and the combined screener and main interview response rate (R_{at}).

Each quarter, the expected number of hours to work was based on the labor model specified earlier. The target that interviewers had for their hours each week is 30. This number was usually used in the sample selection equation. Managers monitored interviewers to ensure that they met the target for hours. The sample line assignment process starts from expected hours, say H_{at} for the α^{th} PSU (usually 360 hours per interviewer) at calendar quarter *t.* A unique estimate of the HPI, HPI_{at}, is generated for each PSU. Estimates for occupancy, eligibility, the subsampling rates, and response rates, $\hat{O}_{at}$, $\hat{E}_{at}$, $\hat{S}_{at}$, and $\hat{R}_{at}$, although denoted at the PSU-level, were actually developed for the sample as a whole. Attempting to estimate these parameters at lower levels (e.g. Census Region) simply led to more variance in the probability of selection weights and did not prove to be accurate. The following formula was estimated for each PSU.

$$A_{at} = \frac{\left(H_{at} / HPI_{at}\right)}{\left(\hat{E}_{at} \times \hat{O}_{at} \times \hat{S}_{at} \times \hat{R}_{at}\right)}$$

For each PSU α during calendar quarter *t*, the ratio of lines needed for an efficient workload over the lines allocated under an EPSEM sample of housing units is defined:

$$D_{at} = \frac{A_{at}}{\sum_{b=1}^{a_a} L_{bt}}$$

This ratio is then used to modify the sample size in each segment for PSU α. Here the notation for PSUs, segments, and time is suppressed for π_3^* and π_3:

$$\pi_3^* = \pi_3 \times D_{\alpha t} \qquad (4)$$

Note that this rate might imply a non-integer value number of sampled housing units. Therefore, the probability of selection was not the number of units selected divided by the number of units on the list. The latter rate is close to the actual rate, but may have been rounded up or down because of the need to select an integer number of housing units. Further, during NSFG 2015-2017, the ratio $D_{\alpha t}$ was bounded (in order to control the variability of the weights) to be no more than 2.5 and no less than 0.5.

In a final step, the rates of selection π_3^* were modified by factors F_m designed to produce the desired sampling rates across the housing unit strata denoted l. The number of strata varied over the quarters. There were three strata (predicted low, medium and high probability of being eligible).

$$\pi_{3m}^{\dagger} = \pi_3^* \times F_m$$

The adjustment factors F_m were set based upon a review of the expected 1+L weighting loss and the expected increase in the eligibility rate under a distribution of options. The stratum with the highest eligibility had the sampling rates for its units raised. The other strata had their sampling rates lowered by a factor that would keep the sample size nearly constant. Given the link between the sampling rates and interviewer productivity, there was a need to implement this change gradually. Therefore, the expected percentage increase in the eligibility rate increases over time. The adjustment factors F_m applied to the sampling rates during 2015-2017 are included in Table 3.

Table 3. Proportion of Housing Units in Each Stratum, Sampling Rate Adjustment Factor, and Predicted Eligibility by Quarter for 2015-2017

Quarter	Predicted Low Eligibility			Predicted Medium Eligibility			Predicted High Eligibility			Expected Percentage Point Increase in Eligibility	Actual Percentage Point Increase in Eligibility
	% In Stratum	Predicted Eligibility	Adj Factor F_m	% In Stratum	Predicted Eligibility	Adj Factor F_m	% In Stratum	Predicted Eligibility	Adj Factor F_m		
Q17	24%	25%	0.78	30%	59%	1.00	46%	86%	1.15	4.2%	4.1%
Q18	16%	17%	0.71	37%	50%	1.00	47%	83%	1.16	4.3%	4.2%
Q19	19%	16%	0.68	27%	46%	1.00	54%	77%	1.18	5.3%	4.9%
Q20	22%	21%	0.69	40%	53%	1.00	38%	82%	1.28	5.5%	6.0%
Q21	21%	23%	0.64	36%	55%	1.00	42%	81%	1.25	5.9%	4.6%
Q22	16%	20%	0.69	32%	50%	0.89	51%	79%	1.25	5.1%	4.8%
Q23	14%	18%	0.67	35%	50%	0.84	50%	80%	1.32	6.0%	5.8%
Q24	15%	18%	0.67	31%	46%	0.79	54%	79%	1.34	6.5%	6.0%

Once the allocation and listing steps had been completed, a sample of housing units was selected systematically from a geographically-sorted list of housing units beginning from a random start using the sampling rates ($\pi_{3m}^{\dagger}$) described in this section.

6 Within-Household Selection

The last stage of sample selection was conducted within the household during the screening activities. An adult member of the household was asked to provide a list of all persons living in the household. Information on the gender and age of each person was recorded in the household screener, and if the household member was within the NSFG age range of 15-49, then information on race and Hispanic origin was collected. Once all household members were covered, interviewers asked additional questions to be sure no one was missed, particularly college students living away from home at a dormitory, fraternity, or sorority. (College students living away from home in their own apartment or housing unit are covered by the household frame, and are not considered to be part of their parents' household.) Dormitory, fraternity, or sorority residents were included in the household listing of their parents' household.

The range of eligible ages was expanded beginning in 2015 from 15-44 to 15-49. This expansion required some changes to within-household selection rates (see Sample Design Documentation for details). The main change was that the expected proportion of interviews to be with teens was reduced from 20.0% to 18.2%. If no one in the household was between the ages of 15 and 49 years, the screening interview concluded with the interviewer thanking the screener informant for his/her time. If the household included one or more age-eligible persons, the computer-assisted screening system made a selection of one eligible person in the household. No more than one eligible person was selected within each household.

Within-household sampling rates for eligible persons varied by age and gender in order to meet the target sample sizes for teens and females. The within-household selection procedure assigned a "measure of size" to each age-eligible person in the household based on the age and sex of the listed person. Larger measures assigned to a subgroup increased the chances that persons in that subgroup were selected for interviewing (see Table 4). Larger measures of size were assigned to teenagers 15-19 years of age in order to select enough to meet sample size targets. Slightly larger measures were also assigned to females to increase the number of females relative to males in the final sample.

Extreme probabilities of selection can result from this algorithm in two situations. The first situation is if there were a large number of persons within a household. These extreme probabilities of selection would always occur for large households under any sample design where one person per household is selected, although the problem may be magnified by the unequal probabilities assigned for the NSFG. The second situation that can result in extreme weights occurs when a person with a low measure of size lives with other persons with larger measures. For example, a 20-49 year old male who lives with three male teens would have ([0.23]/[0.23+3*0.91]=) 0.078 probability of being selected. This would result in a weighting factor of about 12.87 for such a person.

Table 4. Measures of Size for Determining Within-Household Probability of Selection: NSFG 2011-2017

Data Collection Years	Female			Male		
	15-19	20-44	45-49	15-19	20-44	45-49
2011-2015	1.00	0.40	NA	0.93	0.36	NA
2015-2017	1.00	0.25	0.25	0.91	0.23	0.23

If $MOS_{p,i}$ is used to denote the measure of size for the p^{th} person in the i^{th} household with P_i total persons in the i^{th} household, then the following equation can be used to calculate each person's probability of selection:

$$\pi_4 = \frac{MOS_{p,i}}{\sum_{j=1}^{P_i} MOS_{j,i}} \qquad (5)$$

Extreme weights within households occur most often when teens are present in the household. The teens are given high probabilities of selection relative to the adults (approximately a 2.5:1 ratio, see Table 4). This leads to reduced variation in weights among teens, but increases the variation for adults.

In order to reduce this variation, the key statistics were compared for households with adults and teens where an adult was selected for the following two groups:

- Those with a selected adult living with one teen (n=429), and
- Those with a selected adult living with two or more teens (n=60).

The latter group had very high within-household selection weighting factors compared to the former group. The results of this comparison showed that among 18 key statistics, only one was significantly different across the two groups. Even though these comparisons are limited by small sample sizes, any inferences that can be made from this dataset about this subgroup (adults living with teens) will face the same constraint. Therefore, the weights were smoothed across these two groups of adults with teens. The sum of the within-household selection weights was held constant, but the weights for adults living with more than one teen were reduced and the weights for adults living with just one teen were increased by the same total amount, such that the sum of the within-household selection weights remained the same across all 489 cases before and after the adjustment. The smoothed version of this probability of selection is denoted π_4^*.

7 Two-Phase Sampling

Each NSFG calendar quarter consists of two phases. In the first 10 weeks of the quarter, interviewers screened selected lines in assigned segments, conducted main interviews in households with eligible persons, and updated or prepared "scratch" listings for the segments allocated in the next calendar quarter. After 10 weeks of data collection, there remained addresses that had not been successfully screened and sample persons who had not yet completed the interview. If the data collection were halted at the end of 10 weeks, these unscreened lines and sample persons not yet interviewed would contribute to nonresponse bias. A "double or two-phase sample design" (Hansen and Hurwitz, 1946) was instituted for the remaining two weeks of the quarter as a method to reduce the nonresponse bias in survey statistics. Groves and Heeringa (2006) expanded on this design by recommending that design phases include protocols that are complementary such that the biases across the phases are cancelled out.

In the NSFG continuous design for the years 2015-2017, as in prior years of the continuous NSFG, a subsample of one-third of nonrespondents was chosen for weeks 11 and 12 of each quarter based on study of the history of the first 10 weeks' sample. Study staff developed response propensity models to predict the probability that a given case yields a completed screening interview or a completed main interview (see Groves et al., 2009, for details of the propensity models). Within a PSU, two of the three segments were randomly sampled to be included in the second phase. These segments were selected using the sum of the estimated response propensities in each segment as a measure of size and selecting two of three using PPS sampling.

The active nonresponse cases in the two sampled segments were grouped into four strata at the conclusion of the 10-week Phase 1 data collection. The cases were first categorized as unscreened or identified eligible persons. Among the unscreened cases, a person who was predicted to be eligible (based upon a logistic regression model including paradata and sampling frame data used in response propensity models, supplemented with information from commercial databases regarding the ages of persons within unscreened households) were reclassified as an "identified eligible person." Within each of these groups, cases were classified as high or low propensity to respond. This created a 2 x 2 classification of all active cases. A disproportionately allocated sample of nonresponse cases was selected across these groups or second phase strata, with higher probabilities of selection from strata with higher probabilities of response and from strata with known or predicted "eligible persons." These selected lines and persons were then released to interviewers for Phase 2 data collection in the last two weeks of the calendar quarter.

Given this design, the second phase probability of selection of the i^{th} housing unit is a complex function of its screening status and estimated response and eligibility propensities. This probability is denoted as follows:

$$\pi_5 = \frac{\frac{2 \times D_{h\alpha t}}{3} \times MOS_{h\alpha\kappa}}{\sum_{i=1}^{D_{h\alpha t}} MOS_{h\alpha i}} \times \frac{MOS_{h\alpha\kappa\iota}}{\sum_{j=1}^{L_{h\alpha\kappa}} MOS_{h\alpha\kappa j}} \qquad (6),$$

Where $D_{h\alpha t}$ is the number of segments in the h^{th} stratum and the α^{th} PSU during the t^{th} quarter. The $MOS_{h\alpha\kappa}$ is the measure of size for the κ^{th} segment. Note: this measure of size is different than that used in the original selection of the segment. As described above, this MOS is based upon estimated response propensities. $MOS_{h\alpha\kappa\iota}$ is the "measure of size" for the housing unit. Housing units received different measures based on their screening status, expected eligibility for unscreened cases, and probability of response. For cases that were completed during Phase 1, this probability was set to 1.0.

8 Weighting to Compensate for Unequal Probabilities of Selection

The probability of selection of each sample person can thus be computed using the probabilities of selection for PSUs, segments, sample line, within household selection, and Phase 2 subsampling of active cases. Using the notation above, a final sample selection probability can be calculated as:

$$\pi_{h\alpha\beta i\delta\gamma,yr5-6} = \pi_1 \times \pi_{1,yr5-6} \times \pi_2 \times \pi_{2,yr5-6} \times \pi_{3m}^{\dagger} \times \pi_4^{*} \times \pi_5$$

The notation has dropped the indices, as was done above, on the right-hand size, but uses the indices on the left-hand size to emphasize that each unit has a specific weight. The notation denotes the α^{th} PSU in PSU stratum *h,* the β^{th} segment, the i^{th} household, $(\delta)^{th}$ person within the $(h\alpha\beta i)^{th}$ household in a segment, and the phase (γ, either 1 or 2) for the sample selected in Years 5 and 6. The right-hand side stages of selection probabilities have all been defined in the previous sections. The base weight compensating for unequal chances of selection for the $(h\alpha\beta i\delta\gamma)^{th}$ eligible person is the inverse of this probability of selection, $w_{h\alpha\beta i\delta\gamma.yr5-6} = \pi_{h\alpha\beta i\delta\gamma,yr5-6}^{-1}$.

9 Post-Survey Adjustment

9.1 Post-Survey Adjustments for Unit Nonresponse

Both sample-based weighting adjustments and population-based (post-stratification) adjustments were used to reduce error from unit nonresponse for the NSFG 2015-2017. Unit nonresponse is the failure to obtain data for a selected unit by the end of data collection activities. Survey statisticians advocate the use of two kinds of data as nonresponse predictors in the adjustment process. One is paradata collected routinely throughout the data collection process, such as contact observations and call records. The other is deliberate interviewer observation on a limited set of potential weighting adjustment predictors that can be used to develop models more predictive of survey cooperation processes and, simultaneously, the survey data themselves. NSFG 2006-2010, for example, used both of these kinds of data in the unit nonresponse adjustment process (see Lepkowski, et al., 2013). The current data collection includes paradata collection at the listing, the calling, the contact, and the interviewing phases of NSFG, and further development of the collection of interviewer observations on factors thought to be related to nonresponse and to the underlying measures collected in NSFG.

Unit nonresponse occurs in NSFG 2015-2017 at two levels: screening to identify sample eligible persons in sample households and main interviewing among selected eligible persons. There was also nonresponse at the initial contact level in the screener interviewing process, but there is so little data available for non-contact addresses, and so little nonresponse due to noncontact, that a separate adjustment was not feasible. In the following, nonresponse due to a failure to contact sampled households is part of the screener nonresponse. “Sample based” unit nonresponse adjustments were developed by generating predicted probabilities of response using all available data for respondents and nonrespondents at the screener and main interview levels. As noted above, screener and main interview cases have different cooperation processes that call for separate modelling in the adjustment process. In addition, there is slightly different data available at each level. Main interview nonresponse occurs at any time after the conclusion of screening – that is, after a sample person had been selected. The main interview response and nonresponse cases therefore have household composition information including race or ethnicity, age, and sex for all persons in the household. A two-step screener followed by main nonresponse adjustment affords the use of a broad range of sampling frame data and paradata at the screener level adjustment, and the same data plus household composition data for the main interview nonresponse adjustment.

This nonresponse adjustment for the NSFG implements an assumption widely used in the adjustment of survey data – the missing at random (MAR) assumption. A nonresponse weighting adjustment developed under this assumption is computed as the inverse of an estimated response rate or propensity within a subgroup. This is a sample based weight adjustment that, under the MAR assumption, substitutes an estimated response propensity for the probability that a unit will participate in the survey. Thus, the inverse of the predicted probability of response serves as an adjustment factor.

Let $S_i = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$ be a zero-one indicator variable denoting whether a sample address has been successfully screened to determine whether eligible persons lived in the household. The value 1 denotes successful screening and 0 denotes non-contact as well as addresses where screening interviews were refused or not completed for other reasons. This indicator S_i is not defined for unoccupied sample addresses. The screener level logistic regression model is $\pi_{(s)i} = \Pr\left(S_i = 1 \middle| X_{(s)i}\right) = (1 + \exp(-X'_{(s)i}\beta_s))^{-1}$ where $X_{(s)i}$ is a vector of predictor values for the i^{th} occupied housing unit and $\beta_{(s)}$ is a vector of coefficients. Standard maximum likelihood estimation was used to obtain estimated coefficient values $\hat{\beta}_{(s)}$. These in turn were used to predict the probability of screener completion propensity $\hat{\pi}_{(s)i} = \exp\left(-\hat{\lambda}_{(s)i}\right) \Big/ \left(1 + \exp\left(-\hat{\lambda}_{(s)i}\right)\right)$, where $\hat{\lambda}_{(s)i} = X'_{(s)i}\hat{\beta}_s$ is the predicted logit.

At the main interviewing level of adjustment, $M_i = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$ denotes another zero-one main interviewing indicator for the i^{th} successfully screened occupied housing unit. M_i is thus 1 when a selected eligible person has a completed interview, and 0 otherwise. M_i is missing, or undefined, for all sample addresses that were not occupied or a completed screener was not obtained. The main interview logistic regression model will then be

$\pi_{(m)i} = \Pr(M_i = 1 | S_i = 1, X_{(s+m)i}) = (1 + \exp(-X'_{(s+m)i}\beta_{s+m}))^{-1}$ where $X_{(s+m)i}$ is a vector of predictor values for the i^{th} selected eligible person that includes screener as well as household roster data obtained prior to the main interview. Here, $\beta_{(s+m)}$ is a vector of coefficients. Maximum likelihood estimation methods were used to generate $\hat{\beta}_{(s+m)}$ and predicted logits $\hat{\lambda}_{(m)i} = X'_{s+m}\beta_{s+m}$ From the predicted logits, the predicted probability of main interviewing were calculated as $\hat{\pi}_{(m)i} = \exp\left(-\hat{\lambda}_{(m)i}\right) / \left(1 + \exp\left(-\hat{\lambda}_{(m)i}\right)\right)$. The main interviewing unit nonresponse adjustment is thus conditional on having completed a screener interview.

The predictors in $X_{(s)i}$ and $X_{(s+m)i}$ available include the following:

1) counts and rates for the segment from which the housing unit is selected derived from either the 2010 Census or American Community Survey data for the Census Blocks in the segment or for the Census Block Group or Tract in which the segment is located;

2) data obtained from observations made at the segment and housing unit recorded by the interviewer during the segment updating or scratch listing;

3) contact behavior and statements recorded by the interviewer at each contact with anyone within the housing unit;

4) operational measures, such as number of calls to a housing unit, number of calls to the sample person, and interviewer response rate derived from available paradata;

5) for the main interview propensity model, data derived from the household roster and other data collected in the screening interview; and

6) a limited set of interviewer judgments made at the screener or main interviewing level that are of characteristics related to response propensity and related to fertility and family-related phenomena (for example, whether at the screener level the interviewer believes there is anyone under age 15 in the household).

ISR researchers have been investigating the utility of these measures (Kreuter et al., 2010), including methods for improving them (West, 2010). These "tailored" adjustment variables provide the best prospect for reducing bias and, possibly, variance (Little and Vartivarian, 2005). Although commercially-available data have limited utility for adjustment purposes (West, et al., 2015), these variables are included in the modeling process. Paradata regarding the level of effort applied to each case may be strongly related to response, but are often only weakly related to survey data collected in the main interview (Wagner, et al., 2014).

There are a large number of variables in these sets of predictors and not all can be used in a response propensity model, whether at the screener or the main interview level. The search for the set of candidate predictors began by examination of the relationships of the available variables with the key statistics produced by the NSFG. In order to be included in the modeling process, the candidate predictors needed to have at least some correlation with some of the key statistics. This was determined by examining the average correlation across 18 key statistics.

Once the set of candidate predictors had been identified, a step-wise procedure was used to identify the predictors useful for predicting screener and main unit nonresponse for data collected during 2015-2017. The selected set of variables is listed in Tables 5 and 6.

Table 5. Screener response propensity predictors for nonresponse adjustment models: NSFG

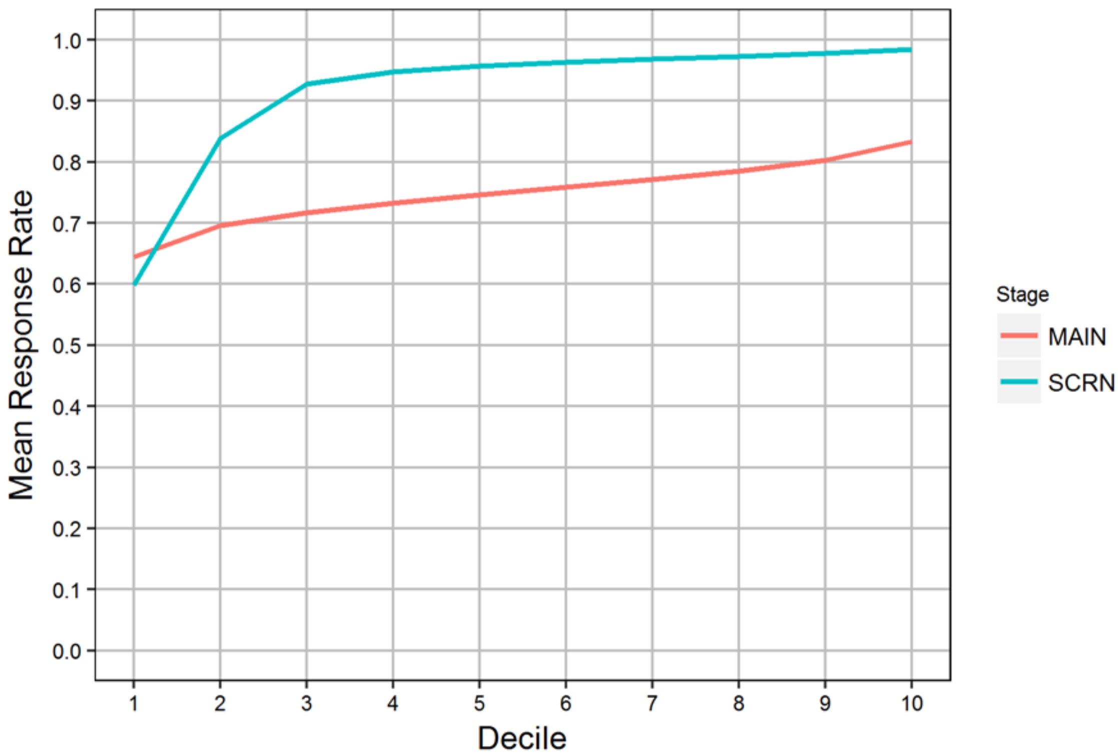
Predictor Name	Description
URBAN	Address in an urban location (yes/no)
SCRN_CALL_CAT	Category for number of Screener Call Attempts (1=1-5; 2=6-8; 3=9+)
MED_HOUSE_VAL_TR_ACS_10_14	Median House Value for the Census Tract Estimated from ACS 2010-2014
AGGREGATE_HH_INC_ACS_10_14	Sum of incomes in the household at the tract level from ACS 2010-2014
RENTER_OCCP_HU_CEN_2010	Number of 2010 Census occupied housing units that are not owner occupied, whether they are renter or occupied without payment or rent at the tract level
BL_HISP_PERC	Percent of population in Census Block Group that is Non-Hispanic Black or Hispanic from Decennial Census 2010.
OCC_RATE	Percent of housing units in the Census Block Group that are occupied from Decennial Census 2010
CHILDRENUNDER15	Interviewer Observation about Presence of Children under the age of 15 in Housing Unit (yes/no)
MSG_NUMBEROFCHILDREN	An estimated count of the number of children in the household. Based upon commercial data. Treated as categorical data with a category for missing.
MSG_AGE	Estimated age of the householder. Based upon commercial data. Treated as categorical data with a category for missing.
MSG_P2AGECALC	Estimated age of “second” adult. Based upon commercial data. Treated as categorical data with a category for missing.
MSG_INCOME	Estimated categorical household income (Low, Medium, High, Missing). Based on commercial data.
MSG_BESTMATCH	A measure provided by the vendor of the likelihood that the match is accurate. A categorical variable with a category for missing.

Table 6. Main-interview, nonresponse-propensity model predictors: NSFG

Predictor name	Predictor description
MED_HOUSE_VAL_TR_ACS_10_14	Median housing unit value for the Census Tract estimated from ACS 2010-2014
BL_HISP_PERC	Percent of population in Census Block Group that is Non-Hispanic Black or Hispanic from Decennial Census 2010.
MANYUNITS	Interviewer observed characteristic of housing units. Categories include single family home vs multi-unit dwelling.
CHILDRENUNDER15	Interviewer Observation about Presence of Children under the age of 15 in Housing Unit (yes/no)
SEXACTIVE	Contact Obs: Respondent Sex Active (yes/no)
SCR_HISP	Screener interview data indicate selected response is Hispanic (yes/no)
SCR_RACE	Screener interview data indicate selected respondent is Black (yes/no)
SCR_AGE_CAT	Screener interview data on age of selected respondent. Divided into four categories (1= 15-19; 2= 20-29; 3= 30-39; 4= 40-49)
SCR_SEX	Screener interviewer data indicated selected respondent is female (yes/no).
SCR_SINGLEHH	Screener interviewer data indicated selected respondent is only adult in the household (yes/no).
WITHIN_HHPROB	The within-household selection probability of the selected respondent.
MSG_DATA_AVAILABLE	Indicates whether commercial data were successfully merged to the housing unit (yes/no).

The use of the inverse of predicted probabilities as unit nonresponse adjustment weights can lead to substantial variation in response propensity weights. A common practice in survey estimation, known as response-propensity stratification, is to reduce this variation by grouping predicted values into classes, and then using a middle value to represent the entire group's predicted values. Since the propensities are estimates, this approach is also more robust to model specification and estimation error. In this case, deciles of the estimated propensity (for both the screener and main models) were created, and each decile was assigned the inverse of the response rate for that decile as a nonresponse adjustment weight at both the screener and main stages. Figure 1 shows the distribution of mean response rates for the screener and main response propensity strata.

Figure 1. Mean Response Rate for the Screener and Main by Deciles of the Propensity



In the final step in the construction of the nonresponse adjustment weight, the distribution of the weights $w_{(r)i}$ was examined for outlying values. The method for selecting predictors, and the use of response-propensity stratification to smooth the weights produced a set of nonresponse adjustments that are limited. The final combined screener and main nonresponse adjustment factors ranged from a low of 1.17 to a high of just under 2.56. The $1+L=\frac{n\sum_{i\in S}w_{4i}^2}{\left(\sum_{i\in S}w_{4i}\right)^2}$ associated with this weighting factor is a modest 1.03. These nonresponse adjustments are mildly associated with some of the key statistics. These associations further limit variance inflation due to weights.

The final step in the construction of the nonresponse adjustment weight was the adjustment of the base sampling weight by the unit nonresponse adjustment weight: $w_{2i}=w_{si}\times w_{(r)i}$, where W_{si} is the base sampling weight.

9.2 Post-stratification

Post-stratification is a population based weighting adjustment. Post-stratification adjustment reduces variances through external population totals for ratio adjustments. These adjustments also reduce bias for noncoverage and nonresponse. Post-stratification has been consistently applied at the last stage of weighting adjustments in the NSFG since Cycle 1 (1973).

Post-stratification is limited to a set of respondent variables on which population estimates are available. Post-stratification by age, gender, and race/ethnicity, is common because of the availability of population estimates of the sizes of those subpopulations from Census and Current Population Survey (CPS) analysis. These were the factors used in the current post-stratification scheme. Let W_g denote the sum of the population in the *g*-th subpopulation and w_g denote the corresponding sum of fifth-step nonresponse adjusted weights for interviewed persons. The simple post-stratification adjustment is the ratio W_g / w_g for each cell.

The control totals are supplied by the U.S. Census Bureau. In previous years, counts of civilian, noninstitutionalized persons were obtained from the Census Bureau. These counts were combined with counts of military personnel not living in group quarters obtained from another source, such as the Defense Manpower Data Center. Starting in 2011-2013 and including2015-2017 NSFG, counts of the household population were obtained from the Census Bureau that also included military personnel in households. These personnel are usually excluded from household populations, but were included here since they are an important segment of the age-eligible population. Unlike counts of civilian, noninstitutionalized persons from the Census Bureau, the counts of the household population used since 2011-2013 do not include individuals living in noninstitutional group quarters, such as college dormitories.

Since data were collected from September 2015 through September 2017, it was necessary to determine a point in time to use as the reference point for the population. The U.S. Census Bureau provided control totals as of July 1 in each year, therefore July 1, 2016 was selected as the reference point for the population, roughly the midpoint of this data collection period. The selected factors used for post-stratification were age (in seven categories: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49), sex, and race/ethnicity (in three categories: black non-Hispanic, non-black non-Hispanic, and Hispanic). This created 42 (7x2x3) separate cells for which population counts were compared to estimated totals. The estimates were based on weights which are the product of the base probability of selection weight and the nonresponse adjustments described earlier. Table 7 presents, for each of the 42 cells, the number of respondents, the population total, the weighted (using selection and nonresponse weights) sample estimate, and the post-stratification adjustment (*W_g /w_g)*.

Table 7. Control Totals, Estimated Totals, and Post-stratification Adjustments used for NSFG 2015-2017

Sex	Race/Ethnicity	Age	Respondent Count	Census Population Totals	NSFG Estimated Total	Post-stratification Adjustment Factor
Male	Black	15-19	201	1,537,844	1,106,486	1.390
		20-24	148	1,580,010	1,414,037	1.117
		25-29	160	1,549,764	1,237,575	1.252
		30-34	123	1,270,525	839,726	1.513
		35-39	106	1,197,999	914,795	1.310
		40-44	84	1,095,492	668,572	1.639
		45-49	116	1,138,846	861,125	1.323
	Hispanic	15-19	260	2,359,463	1,785,969	1.321
		20-24	133	2,349,901	1,838,916	1.278
		25-29	150	2,322,152	1,707,400	1.360
		30-34	131	2,250,684	1,079,053	2.086
		35-39	113	2,151,936	1,359,294	1.583
		40-44	110	1,953,704	1,435,373	1.361
		45-49	89	1,767,256	877,476	2.014
	Other	15-19	425	6,029,356	4,278,417	1.409
		20-24	359	6,453,767	4,597,901	1.404
		25-29	386	7,284,818	4,282,271	1.701
		30-34	415	7,057,332	5,027,501	1.404
		35-39	372	6,697,203	5,163,824	1.297
		40-44	323	6,411,190	5,043,726	1.271
		45-49	336	7,158,002	5,126,987	1.396
Female	Black	15-19	201	1,519,880	1,403,092	1.083
		20-24	214	1,654,845	1,398,284	1.183
		25-29	260	1,723,537	1,163,463	1.481
		30-34	211	1,508,048	1,316,548	1.145
		35-39	156	1,451,485	1,141,798	1.271
		40-44	146	1,350,570	903,517	1.495
		45-49	153	1,397,935	782,262	1.787
	Hispanic	15-19	253	2,270,479	1,654,078	1.373
		20-24	156	2,263,325	1,399,594	1.617
		25-29	207	2,183,216	1,753,022	1.245
		30-34	195	2,113,468	1,521,346	1.389
		35-39	163	2,068,823	1,625,337	1.273
		40-44	128	1,942,468	1,706,623	1.138
		45-49	116	1,755,392	1,432,254	1.226
	Other	15-19	470	5,663,957	5,051,697	1.121
		20-24	374	6,258,438	4,433,006	1.412
		25-29	467	7,250,571	5,052,884	1.435
		30-34	473	7,123,702	5,573,135	1.278
		35-39	433	6,813,279	5,110,962	1.333
		40-44	384	6,559,123	6,400,369	1.025
		45-49	394	7,345,545	5,223,885	1.406

These post-stratification adjustment factors are in line with those from previous cycles of the NSFG. The largest adjustment factor is for Hispanic males 30-34 years of age (2.086), while the smallest factor is for other females 40-44 years of age (1.025).

9.3 Weight Trimming

Extreme variation in weights can inflate the variance of survey estimates. Often, it is the case that the most extreme weights can inflate the variance while producing only trivial changes in the estimates. In this situation, the extreme weights only inflate the total mean squared error. Trimming these weights is a common practice for surveys in order to reduce the estimated variance without increasing any nonresponse bias. Considerable reduction of the variability of the weights can be achieved by a reduction of a few extremely large weights. Reduction of variation in the weights was achieved by smoothing some differences in weights as described in previous sections. For instance, the weights of adults living with teens were smoothed.

The weight trimming process took the following steps. First, the variation in the weights was examined. Outlying weights at both ends of the distribution (i.e., very small and very large weights) were identified. Table 8 shows percentiles at the high and low tails of the distribution of the final, untrimmed weight.

Table 8. Percentiles of the Untrimmed Weight (Including probability of selection, nonresponse adjustment, and post-stratification weighting factors): NSFG 2015-2017

Percentile	Weight	Percentile	Weight
0 (min)	1,300	100 (max)	695,533
1	1,920	99	106,744
2	2,155	98	72,584
3	2,136	97	59,501
4	2,489	96	50,751
5	2,588	95	44,820
6	2,705	94	40,375
7	2,842	93	36,770
8	2,951	92	33,716
9	3,042	91	31,367
10	3,142	90	29,354

The impact on estimates of trimming the tails of this distribution was then examined. The trimming included taking the sum of the trimmed weights within each post-stratification cell, and redistributing it proportionately across the cases that were not trimmed within the same cell. This was done iteratively until no weight was above the specified minimum or maximum value for the weights. This had the effect of maintaining the post-stratification after the trimming step was complete. This step was completed with different levels of trimming. For each level of trimming, the impact on point estimates and variances across several key statistics was evaluated. Trials of trimming of the following percentiles were made: The 1st and 99th percentiles, 2nd and 98th, 3rd and 97th, 5th and 95th and 10th and 90th percentiles. The trimmed weights were then used to estimate the 19 key statistics (9 for females and 10 for males). The criterion for selecting which weights to trim could be reduction in Root Mean Squared Error (RMSE). However, since this was evaluated only for a subset of key estimates, a somewhat conservative level of trimming was chosen, rather than risking introducing bias into estimated quantities that were not considered in the evaluation. The decision was made to trim the weights at the 1st and 99th percentiles. Table 9 shows the estimated means and variances of key statistics using weights that have had the "tails" trimmed at various percentiles. In this table, the "bias" of an estimate is derived as the difference between the estimate using the full, untrimmed weights, and estimates that use weights that have had the extreme weights trimmed. This table includes the results for two different trimmed weights.

Table 9. Estimates of Mean and Variance for Key Statistics Using Full Weight and Two Trimmed Weights

		Full Wgt (No Trim)		Trim: 1% and 99%				Trim: 2% and 98%			
Var Name	Description	Mean	Var x 1,000	Mean	Var x 1,000	"Bias" x 1,000	RMSE x 1,000	Mean	Var x 1,000	"Bias" x 1,000	RMSE x 1,000
CMLSXP	Ever Had Sex in Past 12 Months	0.909	0.059	0.909	0.044	-0.274	6.622	0.908	0.039	-0.995	6.305
CRALL	Number of Co-residential Children	0.909	1.339	0.892	1.032	-17.616	36.641	0.880	0.891	-29.452	41.938
EVRCOHAB	Ever Cohabiting	0.319	0.145	0.321	0.113	2.022	10.816	0.322	0.101	2.880	10.431
FEMCMLASTSEX	Ever Had Sex in Past 12 Months	0.896	0.054	0.892	0.055	-3.910	8.388	0.891	0.051	-5.028	8.731
FEMEVERPREG	Ever Pregnant	0.643	0.142	0.641	0.106	-2.036	10.514	0.641	0.092	-1.454	9.718
FEMEVRCOHAB	Ever Cohabiting	0.564	0.160	0.568	0.113	3.174	11.110	0.567	0.099	2.381	10.237
FEMMARSTAT	Never Married	0.358	0.133	0.365	0.103	6.620	12.103	0.367	0.090	8.937	13.044
FEMMONSX	Months of Non-Intercourse in Past 12 Months	2.732	13.224	2.698	8.245	-33.773	96.880	2.715	7.596	-16.820	88.763
FEMNUMBABES	Number of Live Births	1.280	1.121	1.278	0.848	-2.540	29.236	1.282	0.772	1.088	27.810
FEMPARTS12	Number of Sexual Partners in Past 12 Months	1.162	0.492	1.157	0.454	-4.580	21.800	1.158	0.456	-3.162	21.589
FEMRSTRSTAT	Using Any Method of Sterilization	0.208	0.101	0.210	0.082	2.288	9.336	0.210	0.072	2.012	8.726
LSXUSEP	Using Any Method of Contraception At Last Sex with Last Sexual Partner	0.873	0.216	0.878	0.116	5.076	11.895	0.878	0.103	4.245	10.994
MARSTAT	Never Married	0.428	0.161	0.433	0.122	4.676	11.983	0.437	0.108	8.399	13.360
METHHIST	Using Any Method of Contraception At Last Sex with Last Sexual Partner	0.717	0.156	0.717	0.100	0.520	10.015	0.714	0.091	-2.513	9.847
MON12PRTS	Number of Sexual Partners in Past 12 Months	1.070	0.426	1.072	0.399	2.108	20.075	1.073	0.370	3.638	19.582
NCALL	Number of Non Co-residential Children	0.173	0.164	0.174	0.144	1.334	12.059	0.177	0.143	4.148	12.665
RHADSEX	Ever Had Sex	0.868	0.055	0.865	0.048	-3.094	7.584	0.862	0.044	-5.398	8.576
ROSCNT	Number of Household Residents	3.623	1.255	3.612	1.078	-10.815	34.570	3.601	0.952	-21.595	37.662
RSTRSTAT	Using Any Method of Sterilization	0.099	0.119	0.093	0.044	-5.906	8.899	0.094	0.041	-5.095	8.181
TOTAL RMSE							0.371				0.368
MEAN RMSE							0.020				0.019

After trimming the extreme weights, the cases that had not been trimmed had their weights increased such that the sum of the weights within each cell was still equal to the population control total. If N_g denotes the population count in cell *g* , the set of cases that are trimmed are denoted T , and the weight for case *i* in group *g* after trimming is denoted w_{gi} , then the procedure was to reweight the cases that were not trimmed within the cell to equal the population count minus the sum of the trimmed weights. This is done by multiplying each weight in cell *g* that was not trimmed by a constant: $k_g = \frac{N_g - \sum_{i \in T} w_{gi}}{\sum_{i \notin T} w_{gi}}$. If any weight was increased above the specified level for trimming the weights, the trimming and re-post-stratification steps were repeated until no weight exceeded the specified limits.

10 Final Weight

The final weight is a combination of the procedures described in the report for developing the selection weight, nonresponse adjustments, and post-stratification adjustment factors. The trimmed weight includes an additional trimming and post-stratification procedure. Variance estimation examples using the two-year weight for 2015-2017 can be found on the web page "2015-2017 NSFG: Public-Use Data Files, Codebooks, and Documentation" under "Variance Estimation Examples". A four-year weight was also created to enable analyses that combine data from 2015-2017 with data from 2013-2015. These four-year weights reflect the probabilities of selection for units over that four-year interval, nonresponse adjustments, and post-stratification factors. The four-year weights were post-stratified to estimated population totals as of July 1, 2015. A six-year weight was also created for use with combined data from 2011-2013, 2013-2015, and 2015-2017. These six-year weights reflect the probabilities of selection for units over that six-year interval, nonresponse adjustments, and post-stratification factors for that time period. The six-year weights werepost-stratified to estimated population totals as of July 1, 2014. See the 2015-2017 NSFG User's Guide, Appendix 2, SAS and STATA Syntax Guidelines for Combining Data Across File Releases for program statements in SAS and Stata and guidance for combining NSFG data across survey years and applying these weights. Table 10 lists each variable and gives a short description.

Table 10. Weight Variable Descriptions

Variable	Description
CASEID	Respondent ID Number
WGT2015_2017	**Final Weight Variable** for the **two-year** dataset. Includes the following components: a base selection weight, nonresponse adjustment, and a post-stratification adjustment factor. This weight has had extreme values trimmed and then been re-post-stratified to population control totals.
WGT2013_2017	**Final Weight Variable** for the **four-year** dataset. Includes the following components: a base selection weight, nonresponse adjustment, and a post-stratification adjustment factor. This weight has had extreme values trimmed and then been re-post-stratified to population control totals. Persons 45-49 years (interviewed only in 2015-2017) do not have a value for this weight.
WGT2011_2017	**Final Weight Variable** for the **six-year** dataset. Includes the following components: a base selection weight, nonresponse adjustment, and a post-stratification adjustment factor. This weight has had extreme values trimmed and then been re-post-stratified to population control totals. Persons 45-49 years (interviewed only in 2015-2017) do not have a value for this weight.

For a Glossary of terms used in this document and related documents, see Appendix I in "2015-2017 National Survey of Family Growth (NSFG): Summary of Design and Data Collection Methods"

11 References

Groves, R.M., and Heeringa, S. (2006). "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society, Series A*, 439-457.

Groves, R.M., Mosher, W.D., Lepkowski, J., and Kirgis, N.G. (2009). Planning and Development of the Continuous National Survey of Family Growth. National Center for Health Statistics. *Vital Health Stat 1*(48).

Hansen, M.H., and Hurwitz, W.N. (1946). "The Problem of Nonresponse in Sample Surveys." *Journal of the American Statistical Association*, 41: 517-529.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. (2010). "Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust For Non-Response: Examples from Multiple Surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 173(2): 389-407.

Lepkowski, J.M., Mosher, W.D., Groves, R.M., West, B.T., Wagner, J., and Gu, H. (2013). Responsive Design, Weighting, and Variance Estimation in the 2006-2010 National Survey of Family Growth, National Center for Health Statistics. Vital and Health Statistics Series **2**.

Little, R.J.A., and Vartivarian, S. (2005). "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31(2): 161-168.

Wagner, J., Valliant, R., Hubbard, F., and Jiang, L. (2014). "Level-of-Effort Paradata and Nonresponse Adjustment Models for a National Face-to-Face Survey." *Journal of Survey Statistics and Methodology* **2**(4): 410-432.

West, B.T. (2010). "An Examination of the Quality and Utility of Interviewer Estimates of Household Characteristics in the National Survey of Family Growth." Paper presented at the annual meeting of the American Association for Public Opinion Research, Chicago, May.

West, B.T., Wagner, J., Hubbard, F., and Gu, H. (2015). "The Utility of Alternative Commercial Data Sources for Survey Operations and Estimation: Evidence from the National Survey of Family Growth*." Journal of Survey Statistics and Methodology* **3**(2): 240-264.