

2015-2017 NSFG USER'S GUIDE

Appendix 6: Frequently Asked Questions (FAQ) about the NSFG

List of Questions

1. [How are data collected for the NSFG?](#)
2. [How were people selected?](#)
3. [Why are there 3 different data files for 2015-2017 NSFG?](#)
4. [How do I access the NSFG public use data files?](#)
5. [How do I read the NSFG data files into my statistical software packages?](#)
 - a. [In SAS?](#)
 - b. [In STATA?](#)
 - c. [In SPSS?](#)
6. [Where can I find the codebooks and questionnaires for the 2015-2017 NSFG?](#)
7. [Do I need to use the sampling weights?](#)
8. [Why do you recommend using the recodes?](#)
9. [Where can I find more information on how missing data are handled?](#)
10. [What is the minimum cell size used to create reliable estimates?](#)
11. [How do I find out about skip patterns for a question or questionnaire section?](#)
12. [How do I combine data from different NSFG data files?](#)
13. [How much has the NSFG questionnaire changed since 2013-2015?](#)
14. [Why are certain variables I see in the questionnaire not included \(or not included in the same way\) on the public use file?](#)
15. [Can I identify what region, state, or county a respondent lives in?](#)
16. [Can I analyze the data for just one year, or just one quarter?](#)
17. [Can I analyze the data separately for different states in the country?](#)
18. [Do the variable formats, as used in Webdoc, reflect the actual values in the data?](#)
19. [Can you combine estimates \(e.g., percentages, means\) across NSFG surveys to make one point estimate for a particular time period?](#)
20. [I have other questions about using the NSFG data. Where can I find help?](#)

1) ***How are data collected for the National Survey of Family Growth (NSFG)? How were interviews done and by whom?***

NSFG data included in these public-use files were collected using in-person interviews by trained female interviewers, in respondents' homes. The interviews were conducted from September 2015 to September 2017 with a nationally representative sample of males and females 15-49 years of age. The sample design and questionnaire content of the 2015-2017 NSFG was similar to the 2013-2015 NSFG. At the time of this writing, updated information pertaining to the specifics of the 2015-2017 sample design, weighting, and sample error estimation are in preparation, but the essential details on how the 2015-2017 NSFG was planned and carried out can be found in this documentation posted for the 2013-2015 NSFG:

- [2013-2015 National Survey of Family Growth \(NSFG\): Summary of Design and Data Collection Methods](#)
- [2013-2015 National Survey of Family Growth \(NSFG\): Sample Design Documentation](#)
- [2013-2015 National Survey of Family Growth \(NSFG\): Sample Error Estimation Design](#)
- [2013-2015 National Survey of Family Growth \(NSFG\): Weighting Design Documentation](#)

Additionally, users can refer to these 3 reports that provide more background information about NSFG's move from a periodic to a continuous fieldwork design, as well as information on the use of responsive design strategies to enhance the efficiency of the continuous NSFG:

- *RM Groves et al. Planning and Development of the Continuous National Survey of Family Growth. Vital and Health Statistics Series 1(48). Sept 2009. Hyattsville, MD: National Center for Health Statistics, available at http://www.cdc.gov/nchs/data/series/sr_01/sr01_048.pdf*
- *Lepkowski JM, Mosher WD, Davis KE, et al. The 2006–2010 National Survey of Family Growth: Sample design and analysis of a continuous survey. National Center for Health Statistics. Vital Health Stat 2(150). 2010. Hyattsville, MD: National Center for Health Statistics, available at http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf*
- *JM Lepkowski et al. 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(158). June 2013. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_158.pdf*

2) *How were people selected?*

After a brief screening interview to determine if the household includes any eligible individuals, only one eligible household member is selected for the NSFG interview. Different population subgroups are selected for the survey at varying rates. Hispanics, blacks, and teens were selected at higher rates than others in the 15-49 age group. Women also had a slightly higher probability of selection than men. For example, a teen in a household would have a higher probability of selection than an adult. More information on sampling is available in the NSFG 2015-2017 User Guide Main Text.

3) *Why are there 3 different NSFG data files for 2015-2017?*

As in all NSFG file releases since 2002, there is a separate file for each of 3 types of NSFG data records:

- **The female respondent file** contains 5,554 records - one record **per woman interviewed in 2015-2017**.
- **The female pregnancy file** contains 9,553 records - one record **per pregnancy reported by female respondents interviewed in 2015-2017**. The pregnancy file includes each respondent's complete pregnancy history, regardless of the year in which the pregnancy occurred. If a female respondent has never been pregnant, she has no pregnancy records; if she has been pregnant 5 times, she has 5 pregnancy records in the file. In order to reduce the need for merging files, the respondent file includes variables from the pregnancy file, and the pregnancy file includes selected variables from the respondent file. (Please consult Part 2 of the User's Guide for further information on the quality of reporting for pregnancies, particularly births and abortions, in the NSFG.)

- The **male respondent file** contains 4,540 records - one record **per man interviewed in 2015-2017**. This file includes all information on births and other pregnancies fathered by men in the sample, but this information is presented with the man as the unit of observation; there is no “pregnancy-based” file for male NSFG respondents.

See section on “**Organization of the 2015-2017 NSFG Public Use Data Files**” in Part 1 of the User’s Guide; also see the File Indexes in **Appendix 1** for full lists of all the variables contained in these 3 public-use data files, and Appendix 7 for information on analytic variables suppressed or modified for public use.

4) ***How do I download the NSFG public use data files?***

The female, pregnancy, and male public use data files in ASCII format are available for download on the website.

To download an ASCII data file:

Click on its link from the NSFG webpage. A Data User’s Agreement page will then open that outlines the conditions under which you agree to use these data.

Click on “I Accept These Terms” and an ftp directory will open.

Right-click on the file you want to download, then click on "Save Target As..." (Internet Explorer) or "Save Link As..." (Firefox)

(On a Macintosh, click on the link with the option key held down.)

5) ***How do I read the NSFG data files into my statistical software packages?***

Our website provides SAS, Stata, and SPSS program (or “setup”) statements that assign the name, type, column location, and variable label for each variable in the ASCII data files. The setup files (ending in .SAS, .DO and .DCT, or .SPS) contain instructions on how to read the ASCII data into these statistical software packages. It is important that you read the comments section at the top of the program file, which explains what portions of the program need to be modified to run the program from your computer. Formats in the SAS, STATA, and SPSS program statements are provided for user convenience and ease of display. Users should decide whether using these formats is appropriate for their particular project, or whether they should use different value labels or groupings of categories. See question 18 for more information on formats and value labels.

a) In SAS?

To read the data files into SAS, you will need to download both the ASCII data file (ending in .dat) and the SAS setup file (ending in .sas). When you open up the *.sas file you will see text similar to the text below. You will need to replace the text in bold with the correct location (file path) in which the files are saved on your computer.

```
DATA FEMALE ;
INFILE "C:\Documents and Settings\2015-2017_FemrespData.dat" LRECL=5084 ;
```

b) In STATA?

To read the data files into Stata, you will need to download the ASCII data file (ending in .dat), the do file (ending in .do) and the dictionary file (ending in .dct). When you open up the do file you will see text similar to the text below. You will need to replace the text in bold with the

correct location (file path) in which the files are saved on your computer. In addition, you will need to make sure that the local *.dat file name and the local *.dct file name match. Once the do file is set up, you can run the file and it will generate an output file you can use in Stata (female.dta in the example below).

```
/* The following line should contain the complete path and name of your raw data
file */
local dat_name "C:\Documents and Settings\2015-2017_FemrespData.dat"

/* The following line should contain the path to your output '.dta' file */
local dta_name "C:\Documents and Settings\2015-2017_FemrespData.dta"

/* The following line should contain the path to the data dictionary file */
local dct_name "C:\Documents and Settings\2015-2017_FemRespSetup.dct"

infile "dct_name", using "dat_name" clear
```

c) In SPSS?

To read the data files into SPSS you will need to download both the ASCII data file (ending in .dat) and the SPSS setup file (ending in .sps). When you open up the *.sps setup file you will see text similar to the text below. You will need to replace the text in bold with the correct location (file path) in which the files are saved on your computer. In addition, you will need to make sure that the local *.dat file has the correct file name.

```
FILE HANDLE DATA / NAME= "C:\Documents and Settings\2015-2017_FemrespData.dat" LRECL=5084.
```

6) *Where can I find the codebooks and questionnaires for the 2015-2017 NSFG?*

The 2015-2017 NSFG **codebooks** are accessible on the NSFG webpage via the interactive online “Webdoc” housed on our contractor’s (University of Michigan) website. Please see the User’s Guide section called “**Description of Codebooks**” for further information on using Webdoc, as well as details on the elements of each codebook entry: variable name, variable type, question text, universe statements (“applicable specifications”), response categories and unweighted frequencies, and where relevant, links to recode specifications and special user notes.

The 2015-2017 NSFG **questionnaires** are available on the NSFG webpage in 2 formats – CAPI-lite, an abridged version that shows essential question wording and routing, and the CAPI Reference Questionnaire (CRQ) version that shows full specifications for the interview.

7) *Do I need to use the sampling weights?*

Yes, it is essential to use the sampling weights when analyzing the NSFG data. The NSFG is not based upon simple random sampling, but upon a multi-stage, probability-based

complex sample design, intended to yield estimates representative of the U.S. household population aged 15-49. To control the costs of data collection and to obtain adequate sample sizes, the NSFG sampled some population groups at higher rates than others. As a result of this oversampling and other factors for which the sampling weights are adjusted, such as survey nonresponse, each respondent in the NSFG sample represents a different number of people in the US household population. The number each respondent represents is called a sampling weight. As in previous NSFG file releases, the sampling weight can vary significantly across respondents, so using the sampling weights is critical for producing accurate statistics.

WGT2015_2017 is the final, fully adjusted weight that should be used in all analyses using the NSFG sample of 4,540 male and 5,554 female respondents interviewed over the 24-month period from September 2015-September 2017. In addition to using sampling weights, researchers must use the design variables for the sampling stratum (SEST) and cluster (SECU) to obtain correct standard errors for their estimates.

In addition to the 2015-2017 NSFG public-use files that contain the 2-year sample weight (WGT2015_2017) for the full age range of 15-49, there is a 3-variable mini-file with:

- CASEID
- WGT2013_2017
- WGT2011_2017

The 4-year weight variable WGT2013_2017 corresponds to the 19,095 NSFG cases aged 15-44 interviewed from September 2013 through September 2017 and will allow analyses of combined data from 2013-2015 and 2015-2017 public use files. The 6-year weight variable WGT2011_2017 corresponds to the 29,511 NSFG cases aged 15-44 interviewed from September 2011 through September 2017 and will allow analyses of combined data from 3 public use files: 2011-2013, 2013-2015, and 2015-2017 NSFG. Please note that given the age range expansion to 15-49 in 2015, in order to pool data with previous NSFG data files the 4-year and 6-year weights are limited to respondents aged 15-44 in 2015-2017.

There are several places users can go to get more information about weighting and variance estimation in the NSFG, including:

- The 2013-2015 NSFG User's Guide section on "**Sample Weights and Variance Estimation.**"
- [2013-2015 National Survey of Family Growth \(NSFG\): Sample Error Estimation Design](#)
- [2013-2015 National Survey of Family Growth \(NSFG\): Weighting Design Documentation](#)

8) *Why do you recommend using the recodes?*

Recodes in the NSFG are frequently used "constructed" variables that NCHS has checked for consistency and **imputed** missing values. Many of the variables used in NCHS reports are recodes, and this makes it easier for researchers to replicate NCHS results. While some recodes are straightforward, others represent complex measures and can save you work in your analyses. Imputed recodes are not available for all measures, but to the extent possible, researchers are urged to use the available recodes because they have been checked extensively and their missing values have been handled in a uniform manner. See the

section on **“Recodes and Imputation”** in the User’s Guide for more information. You will find a list of some commonly used recodes in that section. Also, in the codebook documentation and in the File Indexes (**Appendix 1**), recodes are indicated in the “variable type.” Further, in the online Webdoc, if a variable has a recode version, it has a note indicating the recode that corresponds to it. Recode specifications describing how each recode was defined and imputed are provided in **Appendix 3**.

9) *Where can I find more information on how missing data are handled?*

The handling of missing data for most variables is discussed in several sections of the User’s Guide, Part 1, but particularly the sections on **“Coding for ‘Don’t Know,’ ‘Refused,’ and ‘Not Ascertained’ Values”** and **“Century Month Coding for Dates.”** Missing data are imputed for recoded variables (see Question #8).

10) *What is the minimum cell size used to create reliable estimates?*

Percentages should not be shown if the sample denominator is fewer than 100 cases, or the numerator is fewer than 5 cases for reports based on NSFG data. For NCHS reports based on NSFG data, the multistep Standards for Proportions are used to identify whether estimates are reliable, however, these standards are not required for external reports. For more information, please review the NCHS Data Presentation Standards for Proportions https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf

11) *How do I find out about skip patterns for a question or questionnaire section?*

The questionnaires available on the NSFG webpage show skip patterns for every question; see User’s Guide section **“Description of Questionnaires”** for further details. Also, each variable’s codebook entry includes a “universe statement” (also known as applicable specification) that indicates the set of cases for which the variable is asked or defined. These may be as simple as “Applicable for all respondents,” or they may be quite complex. For further information, see **“Universe Statements (“Applicable Specifications”)**” within the User’s Guide section called **“Description of Codebooks.”**

12) *How do I combine data from different NSFG data files?*

Appendix 2 of the User’s Guide, “SAS and STATA Syntax Guidelines for Combining Data Across File Releases,” provides technical guidance and suggested program syntax for combining NSFG data from 2006-2010, 2011-2013, 2013-2015 and 2015-2017.

Given the age range expansion of the NSFG to 15-49 starting in 2015, users are reminded that combining 2015-2017 data with earlier file releases must be limited to respondents aged 15-44. Below are the sample sizes of respondents aged 15-44 when combining data for 2013-2017 and 2011-2017, as also shown in Appendix 2:

Combining data for:	Total N	Males	Females
2013-2017	19,095	8,505	10,590
2011-2017	29,511	13,320	16,191

Before combining data from any NSFG files, however, it is also advisable to determine if the variables you want to analyze are comparable. Two helpful resources for these comparisons are:

- **Appendix 4**, which contains crosswalk grids for all recode variables and their equivalents by sex and across NSFG public-use file releases. These crosswalks show whether comparable recodes exist and whether there are differences in the ways recodes were constructed between males and females or between NSFG public-use file releases.
- **Appendix 5**, which summarizes questionnaire changes between the 2013-2015 and 2015-2017 NSFG.

13) *How has the NSFG questionnaire changed since 2013-2015?*

Most of the questions asked in the NSFG did not change between the 2013-2015 and 2015-2017 surveys. But **Appendix 5** of the User’s Guide contains a summary of questionnaire changes made since 2013-2015 in the 2015-2017 NSFG. We suggest that you consult this appendix to see whether questionnaire changes affect the analyses that you intend to do.

14) *Why are certain variables that I see in the questionnaires not included (or not included in the same way) on the public use file?*

NCHS has a legal and ethical mandate to prevent disclosure of the identities of NSFG respondents. In order to honor that mandate and make as much of the survey data publicly available as possible for researchers, the NSFG staff take a significant set of disclosure-risk-reduction actions before releasing these public use data files, which also undergo intensive review by the NCHS Disclosure Review Board. These risk reduction actions include the suppression or modification of a number of variables that could permit indirect identification of survey respondents, particularly if used in conjunction with other data sources. Please consult the User’s Guide section on “**Protections to Minimize Risk of Disclosure for Individual-Level Data**” for further details. As described there, the codebook entries for those variables that have been modified, and in some cases created, for public use will include a special note. In addition, these variables have been noted with an asterisk in the Public-Use File Indexes provided in **Appendix 1 of the User’s Guide**. Some of the original, full-detail variables that may be of analytic value for researchers are available as restricted-use data through the NCHS Research Data Center (RDC). See **Appendix 7 of the User’s Guide** for further information on these restricted-use analytic variables available to researchers by application through the NCHS RDC.

If you believe that your research will require the use of restricted-use variables, visit the RDC web site (<http://www.cdc.gov/rdc/>). You may also contact the NSFG team at nsfg@cdc.gov if you have questions about using NSFG data in the RDC. There are fees associated with using the RDC

15) *Can I identify what region of the country or what state or county a survey participant lives in?*

NSFG data should not and cannot be used to create state- or lower geographic-level estimates. The survey was designed to create national estimates. State of residence, as well as the 4-category REGION variable (Northeast, Midwest, South, and West), are both suppressed on the NSFG public-use files and will only be available through the NCHS Research Data Center upon approval of a specific research proposal.

16) *Can I analyze the data for just one year, or just one quarter?*

No. Although each year of NSFG fieldwork is designed to be nationally representative of the US household population, sample sizes for a single year are too small to provide estimates with adequate levels of precision. Two years is generally the shortest time frame yielding sufficient numbers of cases and statistical stability. Based on this, weights are not provided for single years of NSFG fieldwork. Please see Appendix 2 for further information on how to appropriately combine data across NSFG file releases to boost sample sizes for analyses. If you have questions about analyzing changes over time by including covariate(s) representing specific years, email nsfg@cdc.gov.

17) *Given that the size of the data set can become large over the quarters of continuous interviewing, can I analyze the data separately for different states in the country?*

No - although the number of data records in the pooled data set of the continuous survey can become large, the sample is limited to a relatively small number of primary sampling areas. These primary areas do not fall in all states in every year, and those that do fall in a given state cannot yield representative estimates for that state. State-level characteristics and other contextual data can be used with the data files in NCHS's Research Data Center. Please see the Research Data Center webpage (www.cdc.gov/rdc) for information on accessing geographic and contextual data.

18) *Does the display of data values in the online codebook reflect the actual values of the variables in the data?*

The values of continuous variables such as numbers of partners, year of an event, or age at an event have sometimes been combined into groups of values in order to shorten the length of the codebook entry, however the separate values are still retained and accessible in the public-use data files. For example, if you see something like "15-19 = 15-19 years" in the

codebook entry, this indicates that separate values of 15 through 19 for this variable are still contained in the data file, but they have been combined for codebook display purposes only. In contrast, if a variable's codebook entry shows "1 = 15-19 years," this means it is a categorical variable where the value 1 represents this group of ages.

19) *Can you combine estimates (e.g., percentages, means) across NSFG surveys to make one point estimate for a particular time period?*

Yes, but there are caveats. Before making a pooled point estimate based on combined data from two or more NSFG survey periods, analysts should test whether there was a significant trend over time for their particular outcome of interest. For example, if the estimates from separate data files vary significantly, the estimate derived from the combined data may be misleading. Similarly, analysts need to use caution in interpreting their point estimate as representing the "average" of the years where data are pooled because: 1) the outcome of interest could have changed over the time period; or 2) the composition of the population may have changed. When combining multiple NSFG data files, analysts should consider their specific analysis goals, define their populations carefully, use provided weight variables, and use caution when interpreting point estimates, particularly population size estimates, based on the combined data files. See Appendix 2 for more detail.

20) *I have other questions about using the NSFG data. Where can I get help?*

The NSFG team has attempted to supply all the information most users will need to use the NSFG data in the various parts of this documentation: the **User's Guide**, the 7 **Appendices**, the **questionnaires**, and the **codebooks**. If you have looked through all of this material and still *cannot find the answer to your question, contact the NSFG team at nsfg@cdc.gov.*