2013-2015 NSFG USER'S GUIDE
Appendix 6: Frequently Asked Questions (FAQ) about the NSFG

List of Questions

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

NSFG data in this public use file were collected using in-person interviews by trained female interviewers, in respondents' homes. The interviews were collected from September 2013 to September 2015 from a nationally representative sample of males and females 15-44 years of age. After a brief screening interview to determine if the household includes any eligible individuals, only 1 eligible household member is selected for the NSFG interview. The sample design and questionnaire content of the 2013-2015 NSFG was similar to the 2011-2013 NSFG. At the time of this writing, updated information pertaining to the specifics of the 2013-2015 sample design, weighting, and sample error estimation are in preparation, but the essential details on how the 2013-2015 NSFG was planned and carried out can be found in this documentation posted for the 2011-2013 NSFG:

- 2011-2013 National Survey of Family Growth: Summary of Design and Data Collection Methods
- 2011-2013 National Survey of Family Growth: Sample Design Documentation
- 2011-2013 National Survey of Family Growth: Sample Error Estimation Design

- 2011-2013 National Survey of Family Growth: Weighting Design Documentation

Additionally, users can refer to these 3 reports which 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) *Why are there 3 different NSFG data files for 2013-2015?*

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 5,699 records - one record **per woman interviewed in 2013-2015**.

- The **female pregnancy file** contains 9,358 records - one record **per pregnancy reported by female respondents interviewed in 2013-2015**. The pregnancy file is intended to include 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. See Appendix 2 for further guidance on merging data from the female respondent and pregnancy files.)

- The **male respondent file** contains 4,506 records - one record **per man interviewed in 2013-2015**. 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 2013-2015 NSFG Public Use Data Files**" in Part 1 of the User's Guide; also see the File Indexes in **Appendices 1a, 1b, and 1c** for full lists of all the variables contained in these 3 data files.

3) *How do I access 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.)

4) *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; however, these formats can be turned off, or commented out, in your statistical package. See question 21 for more information.

5) *Where can I find additional help setting up the data files 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\2013-2015_FemrespData.dat" LRECL=5084;
```

6) *Where can I find additional help setting up the data files 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\2013-2015_FemrespData.dat"
```

/* The following line should contain the path to your output '.dta' file */
local dta_name "**C:\Documents and Settings\2013-2015_FemrespData**.dta"

/* The following line should contain the path to the data dictionary file */
local dct_name "**C:\Documents and Settings\2013-2015_FemRespSetup.dct**"

infile "dct_name", using "dat_name" clear

7) *Where can I find additional help setting up the data files 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**\2013-2015_FemrespData.dat"LRECL=5084.

8) *Where can I find the codebooks and questionnaires for the 2013-2015 NSFG?*

The 2013-2015 NSFG **codebooks** are accessible on the NSFG webpage via the interactive online "Webdoc" housed on our contractor University of Michigan's 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 2013-2015 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.

9) *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 US household population aged 15-44. 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 person in the NSFG sample represents a different number of people in the US household population. The number they represent 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.

WGT2013_2015 is the final, fully adjusted weight that should be used in all analyses using the NSFG sample of 4,506 male and 5,699 female respondents interviewed over the 24-month period from September 2013-September 2015.   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.

As described in **Appendix 2**, there is also a weight variable "WGT2011_2015," made available in a separate file, that should be used when analyzing the combined 2011-2013 and 2013-2015 files when producing point estimates and estimates of population sizes for that four-year period.

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.**"
- 2011-2013 National Survey of Family Growth: Sample Error Estimation Design
- 2011-2013 National Survey of Family Growth: Weighting Design Documentation

10) *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 fairly 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 (page 22). 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.**

11) *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 #10).

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

For NCHS reports based on NSFG data, percentages are not shown if the sample denominator is fewer than 100 cases, or the numerator is fewer than 5 cases. When a percentage or other statistic is not shown for this reason, the table contains an asterisk (*) signifying that the statistic does not meet standards of reliability or precision.

13)   *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."**

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

Appendix 2 of the User's Guide, "SAS and STATA Guidelines for Common File Manipulations," provides technical guidance and suggested program syntax for combining data from:
  -- The female respondent and pregnancy files
  -- The male and female respondent files
  -- Across NSFG data files for females
  -- Across NSFG data files for males

Before combining data from any NSFG files, however, it is 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 data years. 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 cycles.
- **Appendix 5,** which summarizes questionnaire changes between the 2011-2013 and 2013-2015 NSFG.

15)   *How has the NSFG questionnaire changed since 2011-2013?*

Most of the questions in the NSFG questionnaires did not change between the 2011-2013 and 2013-2015 surveys. But **Appendix 5** of the User's Guide contains a summary of questionnaire changes made since 2011-2013 in the 2013-2015 NSFG**.** We suggest that you consult this appendix early to see whether questionnaire changes affect the analyses that you intend to do.

16) *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 mandate to prevent disclosure of the identities of NSFG respondents. In order to honor that mandate and make as much of the survey data available as possible for public use file users, the NCHS Disclosure Review Board required that most geographic information be excluded from NSFG public use data files. In addition, a small number of other survey variables do not appear on the public use files because they potentially increase the risk that a respondent's identity could be discovered. These variables are generally demographic characteristics of household members that may likely be known by others. In most cases, such variables were collapsed or otherwise modified to eliminate or reduce this risk, and these modified versions were included on the public use file. The public use file codebooks will indicate these modified variables with notes such as "This variable has been modified for public use," or "This variable has been top coded for public use."

The original, full-detail variables, as well as geographic information, are available to researchers by application through the NCHS Research Data Center (NCHS/RDC). For a description of these variables available only through the NCHS/RDC, see the User's Guide, **Appendix 7**, **"Variables Modified or Suppressed for Public Use." Appendixes 7a and 7b** provide listings of these restricted-use variables for males and females, and **Appendix 7c** describes the modifications that have been made to the variables in greater detail.
If you believe that your research will require the use of restricted-use variables, visit the RDC web site (http://www.cdc.gov/rdc/) and contact the NSFG team at nsfg@cdc.gov for further guidance. There are fees associated with these requests.

17) *How can I access the ACASI data?*

As done for 2011-2013 NSFG, the data collected in the Audio Computer-Assisted Self Interviewing (ACASI) portion of the interview are included in the main public use files for 2013-2015 NSFG, as released in October 2016. If you wish to access ACASI data from 2006-2010 NSFG or prior NSFG surveys, you must still follow the procedures specified on the NSFG webpage for those datasets.

18) *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 and estimates for the 4 broad Census regions, not state-level or county-level estimates. The respondent's state, county, and region of residence are only available through the NCHS Research Data Center. See question 16 for further information.

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

No. Each year of NSFG fieldwork is designed to be nationally representative of the US household population, and this fieldwork year is divided up into 4 discrete 12-week quarters that cannot be analyzed separately. Sample sizes for a single NSFG 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.

20) *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, 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.

21) *Do the variable formats, as used in Webdoc, reflect the actual values of continuous variables in the data?*

Not necessarily. The majority of the continuous variables in the 3 data files are century month (date) variables, and in all 3 data files these are included with all of their uncollapsed (raw) values, but are shown in Webdoc as groups of years, or ranges of values, to present these data in a more meaningful and concise way.

22) *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, 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; 2) the composition of the population may have changed (or both 1 and 2). One should be cautious when interpreting estimates from combined data files because the NSFG has not been conducted with a continuous nor annual survey design that would permit valid estimation and inference for the full span of years. Additionally, as explained in more detail in

Appendix 2, population sizes based on combined data files should be estimated and interpreted with caution.

23)   *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 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 <u>nsfg@cdc.gov</u>.*