

2011-2013 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. [Why are there 3 different data files for 2011-2013 NSFG?](#)
3. [How do I access the NSFG public use data files?](#)
4. [How do I read the NSFG data files into my statistical software packages?](#)
5. [Where can I find additional help setting up the data files in SAS?](#)
6. [Where can I find additional help setting up the data files in STATA?](#)
7. [Where can I find additional help setting up the data files in SPSS?](#)
8. [Where can I find the codebooks and questionnaires for the NSFG?](#)
9. [Do I need to use the sampling weights?](#)
10. [Why do you recommend using the recodes?](#)
11. [Where can I find more information on how missing data are handled?](#)
12. [What is the minimum cell size used to create reliable estimates?](#)
13. [How do I find out about skip patterns for a question or questionnaire section?](#)
14. [How do I combine data from different NSFG data files?](#)
15. [How much has the NSFG questionnaire changed since 2006-2010?](#)
16. [Why are certain variables in the questionnaires not included on the public use file?](#)
17. [How can I access the ACASI data?](#)
18. [Can I identify what region, state, or county a respondent lives in?](#)
19. [Can I analyze the data for just one year, or just one quarter?](#)
20. [Can I analyze the data separately for different states in the country?](#)
21. [When will the next public use data files be available from the NSFG?](#)
22. [I have questions 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? 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 2011 to September 2013 from a national sample of males and females 15-44 years of age. The design and content of the 2011-2013 NSFG was similar to the 2006-2010 NSFG. For further information on how the NSFG was planned and carried out, see:

- *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.*
- *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 files for 2011-2013?***

There is a separate file for each of 3 types of NSFG data records:

- **The female respondent file** contains one record **per interviewed woman**, and contains 5,601 records.
- **The female pregnancy file** contains one record **per reported pregnancy**, and contains 9,547 records. 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 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.
- **The male respondent file** contains one record **per interviewed man**, and contains 4,815 records. This file includes all information on births and other pregnancies fathered by men in the sample.

See section on “**Organization of the 2011-2013 NSFG Public Use Data Files**” in Part 1 of the User’s Guide; also see **Appendixes 1a, 1b, and 1c** for lists of all the variables contained in these 3 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. Additional advice for users of SAS, Stata, and SPSS is provided in questions 5, 6, and 7.

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. In addition, you will need to make sure that the local *.dat file has the correct file name (see highlighted portion below).

```
DATA ;  
INFILE "C:\Documents and Settings\FemResp.dat" LRECL=4957 ;
```

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 (see the highlighted portion below). 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\FemResp.dat"  
  
/* The following line should contain the path to your output '.dta' file */  
local dta_name "C:\Documents and Settings\female.dta"  
  
/* The following line should contain the path to the data dictionary file */  
local dct_name "C:\Documents and Settings\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 (see highlighted portion below).

```
FILE HANDLE DATA / NAME= "C:\Documents and Settings\FemResp.dat" LRECL=4957.
```

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

The **codebooks** are accessible on the NSFG webpage via the interactive online “Webdoc.” Please see the User’s Guide section called “**Description of Codebooks**” for further information on utilizing 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 **questionnaires** are available on the NSFG webpage in 2 formats as well – 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. Additionally, we have posted detailed outlines for both the male and female questionnaires on the NSFG webpage for a quick reference. See the User’s Guide section called “**Description of Questionnaires**” for further details.

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

Using the sampling weights is critical for producing accurate statistics and correct estimates of the standard errors of those statistics. To control the costs of data collection and to obtain adequate samples, the NSFG sampled some population groups at higher rates than others. As a result, each person in the sample represents a different number of people in the population. The number they represent is called a sampling weight. As in previous NSFG file releases, the sampling weight varies across respondents, so using weighted data is essential for making accurate national estimates from these data.

WGT2011_2013 is the final, fully adjusted weight that should be used in all analyses using the NSFG sample of 10,416 male and female respondents interviewed over the 24-month period from September 2011-September 2013. 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.

See the User's Guide section on "**Sample Weights and Variance Estimation**" for more information.

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," shows how to merge data from the female respondent and pregnancy files using the common CASEID link variable, combine (or pool) data for males and females, and combine (or pool) data across NSFG data years. Before combining any data 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 gender 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 for Year 1 of the 2011-2013 NSFG.

15) *How has the NSFG questionnaire changed since 2006-2010?*

Most of the questions in the NSFG questionnaires did not change between the 2006-2010 survey cycle and 2011-2013. But **Appendix 5** of the User's Guide contains a summary of questionnaire changes in the 2011-2013 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?*

A small number of variables do not appear on the public use file because they increased the risk that a respondent's identity could be discovered. Examples of these are geographic details and characteristics of household members. In most cases, such variables were collapsed or otherwise modified to eliminate this risk, and these versions were retained on the public use file. For a description of these variables, see the User's Guide, **Appendix 7, "Variables Modified or Suppressed for Public Use."** The original, full-detail variables are available to researchers by application through the NCHS Research Data Center (NCHS/RDC). In addition to the description of suppressed and modified variables,

Appendix 7 provides a listing of these restricted-use variables for males and females. 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.

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

In 2006-2010 and prior NSFG file releases, much of the information collected in the Audio Computer-Assisted Self-Interview (ACASI) part of the interview (Female Section J and Male Section K) was not included on the main public use files but instead released separately under a special data use agreement. Unlike past file releases, **the ACASI data for 2011-2013 are now included in the main public use files** released in December 2014. Please see **Questionnaires, Datasets, and Related Documentation** for the 2011-2013 NSFG for all links associated with accessing these data.

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

NCHS has a legal mandate to prevent disclosure of the identities of NSFG respondents. In order to honor that mandate and preserve the detailed marital and fertility histories and demographic information for public use file users, the NCHS Disclosure Review Board required that most geographic information be excluded from NSFG public use data files. Users who wish to use geographic information such as region, state, or county should contact the NSFG team at nsfg@cdc.gov and apply for access through the NCHS Research Data Center.

NSFG data should **not** be used to create state- or lower geographic-level estimates. The survey was designed to create *national* estimates, not state-level estimates.

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

No. Sample sizes for a year are generally too low to provide estimates with adequate levels of precision. Two years is generally the shortest time frame yielding sufficient cases and statistical stability.

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 page 25 of the User's Guide Main Text, or the Research Data Center webpage (www.cdc.gov/rdc), for information on accessing geographic and contextual data.

21) *When will the next public use data files be available from the NSFG?*

We expect that the second set of public use data files from the NSFG, containing interviews conducted in 2013-2015, will be released late 2016. These files will contain data from approximately 10,000 interviews conducted in September 2013 through September 2015.

22) *I have 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 nsfg@cdc.gov.