2006-2010 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 the 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 June 2006 to June 2010 from males and females 15-44 years of age in a national sample of randomly selected households. Only one respondent was selected per household. For further information on how the NSFG was planned and carried out, see these reports:

- o RM Groves et al. Planning and Development of the Continuous National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 48. September 2009. Hyattsville, MD: National Center for Health Statistics. Available at http://www.cdc.gov/nchs/data/series/sr_01/sr01_048.pdf.
- o JM Lepkowski et al. Continuous National Survey of Family Growth: Sample Design, Sampling Weights, Imputation, and Variance Estimation, 2006-2008 *Vital and Health Statistics*, Series 2. No. 150. June 2010. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf

2) *Why are there 3 different data files for 2006-2010 NSFG?*

There is a separate file for each of 3 types of NSFG data records:
- **The female respondent file** contains 12,279 records, with each record representing one **interviewed woman**.
- The **female pregnancy file** contains 20,497 records, with each record representing one **reported pregnancy**. If a female respondent has never been pregnant, she has

no records in the pregnancy file; 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 selected variables from the pregnancy file, and the pregnancy file includes selected variables from the respondent file.

- The **male respondent file** contains 10,403 records, with each record representing one **interviewed man**. This file includes all information on births and other pregnancies fathered by men in the sample.

See section on "**Organization of the 2006-2010 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 data file. 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=5941;
```

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 using "`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=5941 .
```

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

The **codebooks** are accessible on the NSFG webpage in 2 ways – online, interactive "Webdoc" and stand-alone PDF files. Please see the User's Guide section called **"Description of Codebooks"** for further information on these 2 formats, 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. See the User's Guide section called **"Description of Questionnaires"** for further details.

9) ***Do I need to use the sampling weights? Why is there more than one weight on this public use file? Which one should I use?***

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 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, and the weights vary widely.

There are five weights on these files, but for most purposes, WGTQ1Q16 is the one to use. WGTQ1Q16 is the weight for the full sample of 22,682 respondents interviewed over the 4-year period from June 2006-June 2010.

If you are using one of the questions that were added or modified beginning after a specific year of interviewing, then you may want to use one of the other four weights: WGTQ1Q8 (the weight for years 1 and 2 of data collection, quarters 1–8), FINALWGT30 (the weight for the first 2.5 years of data collection, quarters 1–10), WGTQ5Q16 (the weight for years 2, 3, and 4 of data collection, quarters 5–16), or WGTQ9Q16 (the weight for years 3 and 4 of data collection, quarters 9–16). If using selected years of interviewing only, some subgroups may be too small to yield reliable estimates. (See **Appendix 5** for information on additions and changes to questions.) 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 24). Also, in the codebook documentation and in the File Indexes (**Appendix 1**), recodes are indicated in the "variable type." 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)   *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" or "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."**

13)   *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 made for Years 1, 2, and 3 of the Continuous NSFG. No questionnaire changes were made for Year 4.

14)   *How much has the NSFG questionnaire changed since Cycle 6?*

Over 90% of the questions in the NSFG questionnaires did not change between the 2002 survey (Cycle 6) and 2006-2010. **Appendix 5** of the User's Guide contains a summary of questionnaire changes in the 2006-2010 NSFG**.** We suggest that you consult this appendix early to see whether questionnaire changes affect the analyses that you intend to do.

15)   *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 for two main reasons. First, some variables are excluded 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 included on the public use file. For a description of these variables, see the User's Guide, Part 2, **"Variables Modified or Suppressed for Public Use."** The original, full-detail variables are available to researchers through the NCHS Research Data Center (NCHS/RDC). User's Guide **Appendix 7** provides a listing of these restricted-use variables for males and females. If you believe that your research will require restricted-use variables, visit the RDC web site (http://www.cdc.gov/rdc/) and contact the NSFG team at nsfg@cdc.gov.

Second, because of their sensitivity, most variables in the Audio Computer-Assisted Self-Interview (ACASI) part of the interview (Female Section J and Male Section K) are not included on the main public use files, but will be available later upon request. The 2006-2008 ACASI data are currently available; see the 2006-2010 webpage for instructions on how to request these data. A notice will be sent to the NSFG Announcements listserv when the 2006-2010 ACASI data are available. You can subscribe to the NSFG listserv from the NSFG homepage.

Lastly, there were a small number of variables that were not included on the public use files due to data quality issues or programming errors. These are rare occurrences, and are described in Part 2 of the User's Guide. If the variables were retained on the public use files, a note is generally attached to the codebook entry alerting users to the issue.

16) *Can I identify what region of the country or what state or county a respondent lives in?*

No, you cannot do so with the public use data files. NCHS has a legal mandate to prevent disclosure of the identities of NSFG respondents. In order to honor that mandate and preserve the detailed history information and demographic information for public use file users, the NCHS Disclosure Review Board required that geographic information be excluded from NSFG public use data files. Users who wish to use geographic information such as state, county, or region should contact the NSFG team at nsfg@cdc.gov and apply for access through the NCHS Research Data Center, http://www.cdc.gov/rdc/ .

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

The NSFG team has attempted to supply all the information 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.