2017-2019 NSFG USER'S GUIDE
Appendix 6: Frequently Asked Questions (FAQ) about the NSFG

List of Questions

1) *How were data collected for the 2017-2019 National Survey of Family Growth (NSFG)?*

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 2017 to September 2019 with a nationally representative sample of male and female respondents 15-49 years of age. The sample design and questionnaire content of the 2017-2019 NSFG was similar to the 2015-2017 NSFG. Sample design, weighting, and sample error estimation details on how the 2017-2019 NSFG was planned and carried out can be found in this documentation to be posted for the 2017-2019 NSFG, soon after public-use file release (expected by December 2020):

- 2017-2019 National Survey of Family Growth (NSFG): Summary of Design and Data Collection Methods
- 2017-2019 National Survey of Family Growth (NSFG): Sample Design Documentation
- 2017-2019 National Survey of Family Growth (NSFG): Sample Error Estimation Design
- 2017-2019 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 to participate in the NSFG?*

After a brief screening interview to determine if the household included any eligible individuals aged 15-49, only one eligible household member was selected to participate in the survey. Different population subgroups were selected for the survey at varying rates. Teens 15–19 years of age and Black and Hispanic persons were selected at higher rates, yielding an oversample of such persons to ensure large enough sample sizes for reliable statistical analysis. Women also had a slightly higher probability of selection than men. More information on sampling is available in the main text of the User's Guide.

3)   *Why are there 3 different data files for 2017-2019 NSFG?*

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

- The female respondent file contains 6,141 records - one record per woman interviewed in 2017-2019.

- The female pregnancy file contains 10,215 records - one record per pregnancy reported by female respondents interviewed in 2017-2019. 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 file 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 Appendix 2 of the User's Guide for further information on the completeness of women's reporting for live births and abortions in the NSFG.)

- The male respondent file contains 5,206 records - one record per man interviewed in 2017-2019. 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 2017-2019 NSFG Public-Use Data Files**" in the User's Guide main text; also see the File Indexes in **Appendix 1** for full lists of all the variables contained in these three public-use data files, and Appendix 7 for information on analytic variables suppressed, modified, or created 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 project, or whether they should use different value labels or groupings of categories. See question 18 for more information on formats and value labels. Data files are only provided in ASCII format with the accompanying program statements for you to create SAS, Stata, and SPSS datasets. We do not provide SAS, Stata, or SPSS datasets.

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 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\2017-2019_FemrespData.dat" LRECL=3839;
```

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 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**\2017-2019_FemrespData.dat"

/* The following line should contain the path to your output '.dta' file */
local dta_name "**C:\Documents and Settings**\2017-2019_FemrespData.dta"

/* The following line should contain the path to the data dictionary file */
local dct_name "**C:\Documents and Settings**\2017-2019_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 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\2017-2019_FemrespData.dat"LRECL=3839.

6) Where can I find the codebooks and questionnaires for the 2017-2019 NSFG?

The 2017-2019 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 2017-2019 NSFG **questionnaires** are available on the NSFG webpage in two 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? What sampling weights should I use for my analyses?*

Yes, it is essential to use the sampling weights when analyzing the NSFG data. The NSFG uses a multi-stage, probability-based complex sample design, not simple random sampling, 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 U.S. household population. 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.

The two-year sample weight (WGT2017_2019) is the final, fully adjusted weight that should be used in all analyses using the NSFG sample of 5,206 male and 6,141 female respondents interviewed over the 24-month period from September 2017-September 2019.   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.

Additionally, 4-, 6-, and 8-year sample weights for the period 2011-2019 are available in downloadable files on the **2011-2019 Combined Files: Selected Data and Documentation** webpage. Please see this page for further information on combining data files.

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 main text, 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 are the criteria used to assess the reliability of NSFG estimates in Key Statistics and NCHS reports?

Since roughly 2017, NCHS reports have followed the NCHS Data Presentation Standards for Proportions outlined in this report https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf. NSFG data users may find this publication helpful when making decisions about the reliability of estimates produced with NSFG data.

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

The NSFG questionnaires posted on the NSFG webpage are the best way to see the skip patterns or routing for questions in the survey. The questionnaires as provided in two formats: the full specifications laid out in the CAPI Reference Questionnaires (or CRQs) show detailed skip patterns, question wording variants, and consistency checks, while the more streamlined CAPI-lites show an abridged version of skip patterns. See the User's Guide "**Description of Questionnaires"** for further details. In addition to the questionnaires posted on the NSFG webpage, each variable's codebook entry in Webdoc 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?

The **2011-2019 Combined Files: Selected Data and Documentation** webpage provides technical guidance and appropriate weight variables for combining data across the four 2-year file releases from 2011-2019. Weight variables are available for all 4-, 6-, and 8-year file combinations across this data collection period. Guidance is also provided for when users might choose to use the 2-year weights versus the combined-file weights in their analyses with NSFG data from 2011-2019.

Before combining data from any NSFG files from 2011-2019, however, it is also advisable to determine if the variables you want to analyze are comparable. Two helpful resources within the 2017-2019 User's Guide are:

- **Appendix 4**, which contains crosswalk grids for all recode variables and their equivalents by sex and across NSFG public-use file releases in this 8-year data collection period. 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 2015-2017 and 2017-2019 NSFG.

13) *How has the NSFG questionnaire changed since 2015-2017?*

Most of the questions asked in the NSFG did not change between the 2015-2017 and 2017-2019 surveys. **Appendix 5** of the User's Guide contains a summary of questionnaire changes made since 2015-2017 in the 2017-2019 NSFG**.** We suggest that you consult this appendix to see whether questionnaire changes affect your analyses.

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 available publicly as possible for researchers, NSFG staff take a significant set of disclosure-risk-reduction actions before releasing these public-use data files. The files also undergo intensive review by the NCHS Disclosure Review Board. These risk reduction actions include the suppression or modification of several 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 a full listing of these restricted-use analytic variables available through the RDC, as well as further information on the variables that have been modified or created for public use.

If you believe that your research will require the use of restricted-use variables, visit the RDC website for information on the application process. 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 are only available through the NCHS RDC upon approval of a specific research proposal.

The recode variable METRO is included on the public-use file and categorizes the respondent's place of residence at the time of interview in 3 groups: principal city of a

metropolitan statistical area (MSA), other MSA, or not an MSA. These 3 categories can be collapsed to create an indicator of urban-rural residence.

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

No. Although each year of fieldwork is designed to be nationally representative of the U.S. 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 enough numbers of cases and statistical stability. Based on this, weights are not provided for single years of NSFG fieldwork. If you have questions about analyzing changes over time by including covariate(s) representing specific years, email nsfg@cdc.gov.

17) *Can the NSFG be used to provide state-level estimates if data are pooled to increase sample sizes?*

No. The NSFG sample is not designed to yield state-level estimates. Although you may boost the numbers of data records by combining across more than one 2-year data release in 2011-2019, the NSFG sample in any given file release 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 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) *I have reviewed the documentation on the webpage and still have a question. Where can I get help?*

The NSFG team has attempted to supply the information most users will need in the documentation on the webpage: The **User's Guide**, the 7 **Appendices**, the **questionnaires**, and the **codebooks**. If you have reviewed the documentations on the webpage and *cannot find the answer to your question, contact the NSFG team via email at* nsfg@cdc.gov*.*