Public Use Data File Documentation

2013-2015
National Survey of Family Growth

USER'S GUIDE

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
October 2016

Data User's Agreement

Federal law (The Public Health Service Act (Section 308 (d))) provides that these data may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any person or establishment is prohibited by this law.

NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data files. In addition, a small number of records have had one or more responses slightly modified through statistical perturbation. These modifications do not affect point estimates or statistical significance, but are intended to prevent definitive identification of individual respondents.

By using National Survey of Family Growth data, you signify your agreement to comply with the following legal requirements:

1. To use these data for statistical reporting and analysis only;
2. To make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery (301-458-4500); and
3. To not link these data with individually identifiable data from any other data set.

TABLE OF CONTENTS

APPENDICES (provided in separate PDF files posted on NSFG webpage)

Appendix 1: File Indexes for 2013-2015 NSFG Public Use Files
        1a. Female Respondent File Index
        1b. Female Pregnancy (Interval) File Index
        1c. Male File Index

Appendix 2: SAS and STATA Syntax Guidelines for Common File Manipulations
        Combining data from female respondent and female pregnancy files within 2013-2015 NSFG
        Combining data for males and females within 2013-2015 NSFG
        Combining data across NSFG data files

Appendix 3: Recode Specifications for 2013-2015 NSFG
        3a. Female Respondent File Recode Specifications
        3b. Female Pregnancy (Interval) File Recode Specifications
        3c. Male File Recode Specifications

Appendix 4: Recode "Cross-walk" Grids
        4a. Male-Female Recode Crosswalk for 2013-2015
        4b. Female and Pregnancy Recode Crosswalks for 2002, 2006-2010, 2011-2013, and 2013-2015
        4c. Male Recode Crosswalk for 2002, 2006-2010, 2011-2013, 2013-2015

Appendix 5: Summary of NSFG Questionnaire Changes since 2011-2013

Appendix 6: Frequently Asked Questions about the NSFG

Appendix 7: List of Restricted Use Variables Available through the RDC
        7a. List of Restricted-Use Analytic Variables for the Female File
        7b. List of Restricted-Use Analytic Variables for the Male File
        7c. Variables Modified or Suppressed for the Public Use Files

What You Need to Know about This User's Guide

This User's Guide is for researchers planning to use the 2013-2015 National Survey of Family Growth (NSFG) Public Use Data Files. It is intended to guide you through the process of identifying the variables you need for your research and will provide you with essential technical guidance for using the data. The User's Guide consists of:

- **Part 1** providing an overview with information that applies to all users, regardless of their research topics.
- **Part 2** contains information about specific topics or variables relevant to your analysis.
- **7 Appendices** providing further technical documentation and reference material (see page 6 for more detail):

For <u>published reports or statistics based on NSFG data,</u> please see the following:

- The NSFG website (<u>www.cdc.gov/nchs/nsfg.htm</u>) includes lists of **reports and articles using all NSFG file releases**. All NSFG reports published by the National Center for Health Statistics (NCHS) are accessible for download in PDF format from the NSFG website. The cumulative bibliographies posted on the NSFG webpage include NCHS publications, journal articles, and other reports, authored by NSFG staff as well as non-NSFG researchers. NSFG staff attempt to keep these bibliographies up to date. If you notice any omissions, please email nsfg@cdc.gov.
- Another source of published NSFG data is "**Key Statistics**," accessible on the NSFG web page (<u>http://www.cdc.gov/nchs/nsfg/key_statistics.htm</u>). **Key Statistics** provides basic statistics on some of the most commonly cited topics in the NSFG. Within a few months of this public use file release, the Key Statistics page will be updated with 2011-2015 NSFG data.

USER'S GUIDE PART 1: GENERAL INFORMATION FOR USERS OF THE 2013-2015 NSFG PUBLIC USE DATA

As noted on the previous page, the User's Guide for the 2013-2015 NSFG is divided into 2 parts.

Part 1 provides general information for users of the public use data, including:
- An overview of the 2013-2015 NSFG and brief background on previous NSFG surveys
- How to access NSFG data and documentation files
- How the NSFG data files are organized
- Information on sampling weights and variance estimation
- How the NSFG interview data were prepared for public use
- Descriptions of the codebooks and questionnaires

Part 2 ("**topic-specific notes for analysis**") gives information on data quality on selected areas and specific guidance for using particular variables in the data files.

The User's Guide also contains seven appendices providing essential technical documentation and reference information:

1. File indexes (i.e., listings of each variable in the order in which they appear on the file, along with a brief description)
2. SAS & Stata syntax guidelines, as well as some statistical guidance for common file manipulations, including combining data across NSFG file releases
3. Recode specifications for specially constructed and imputed variables
4. Recode cross-walks by sex and across recent NSFG survey years
5. Summary of 2013-2015 NSFG questionnaire changes since 2011-2013
6. Frequently Asked Questions (FAQ) about the NSFG
7. List of variables suppressed or modified for the public-use file including detail about restricted-use variables only available through the NCHS Research Data Center

As noted above, **Appendix 6** of this User's Guide provides a list of **frequently asked questions (FAQ)** about the NSFG, geared toward current or prospective users of the data file. Here are the primary highlights from the FAQ:

1. **Use sample weight (e.g.,WGT2013_2015) and design variables (SECU & SEST)** to make valid estimates. Failure to use the weights and design variables correctly will lead to inaccurate statistical estimates and inferences.

2. **Use recoded variables when available.** They have been edited carefully, and missing

data have been imputed. Some of the most commonly used recodes are listed on page 22.

3. In addition to this 2013-2015 NSFG User's Guide with its seven appendices, the **NSFG website** also provides the following materials:
    a. the **codebooks** (online Webdoc)
    b. the **questionnaires** in two formats
    c. downloadable **ASCII data files**
    d. **program statements** to read in the ASCII data files using SAS, Stata, and SPSS
    e. three **variance estimation examples** using 2013-2015 data which can be adapted for your own research purposes.

4. If you cannot find an answer to your question in the NSFG User's Guide, codebooks, questionnaires, or webpage, please contact NSFG staff at nsfg@cdc.gov or 301-458-4222.

5. The 2013-2015 NSFG sample design, fieldwork plan, and other procedures of the survey are similar to what was used in 2006-2010 and 2011-2013. Until further information is published specific to the survey design and operations of the 2013-2015 NSFG, refer to the section below on "Sample Weights and Variance Estimation," the methodology reports available on the webpage for 2011-2013, as well as the Series 1 and Series 2 technical reports from the 2006-2010 data release available on the NSFG website.

BACKGROUND AND OVERVIEW OF THE 2013-2015 NSFG

The National Survey of Family Growth (NSFG) is designed and administered by the National Center for Health Statistics (NCHS), an agency with the U.S. Department of Health and Human Services' Centers for Disease Control and Prevention (DHHS/CDC). NCHS conducts the NSFG in collaboration with several other agencies of the DHHS. (See Acknowledgments for further detail on these cosponsoring agencies, as well as key personnel at NCHS and at the contractor organization for this file release, the University of Michigan's Institute for Social Research.)

The NSFG became part of the federal statistical system, within NCHS, in 1973. The primary purpose of the survey, particularly since the inclusion of a sample of men as well as women aged 15-44, has been to produce reliable national estimates of:

- factors affecting pregnancy, including sexual activity, contraceptive use, and infertility;
- the medical care associated with contraception, infertility, and childbirth;
- factors affecting marriage, divorce, cohabitation, and family building;
- adoption and caring for non-biological children;
- father involvement with their children;
- use of sexual and reproductive health services; and
- attitudes about sex, childbearing, and marriage.

The survey results are used by the U.S. DHHS and other research and policy organizations to plan health services and health education programs, and to do statistical studies on the topics listed above, among others.

The following table presents basic information on each NSFG public use file release since 1973.

Cycle	Year	Scope	Number of Interviews	Over-Samples	OMB Approved Length	Respondent Incentive	Response Rates
1	1973	Ever-Married Women 15-44	9,797	Black Women	60 Minutes	No	90.2%
2	1976	Ever-Married Women 15-44	8,611	Black Women	60 Minutes	No	82.7%
3	1982	All Women 15-44	7,969	Black Women Teens	60 Minutes	No	79.4%
4	1988	Women 15-44	8,450	Black Women	70 Minutes	No	82.5%
5	1995	Women 15-44	10,847	Black Women Hispanic Women	100 Minutes	$20	78.7%
6	2002	Women 15-44 Men 15-44 (First time)	12,571 W = 7,643 M = 4,928	Blacks, Hispanics, 15-24 year olds	W= 85 min M= 60 min	$40	79% W=80% M=78%
n/a	2006-2010	Women 15-44 Men 15-44	22,682 W = 12,279 M = 10,403	Blacks, Hispanics, Teens	W=80 min M=60 min	$40	77% W=78% M=75%
n/a	2011-2013	Women 15-44 Men 15-44	10,416 W = 5,601 M =4,815	Blacks Hispanics Teens	W=80 min M=60 min	$40	72.8% W=73.4% M=72.1%
n/a	2013-2015	Women 15-44 Men 15-44	10,205 W=5,699 M=4,506	Blacks Hispanics Teens	W=80 min M=60 min	$40	69.3% W=71.2% M=67.1%

NSFG data are based on a multi-stage probability-based, nationally representative sample of the household population aged 15-44. Fieldwork for the 2013-2015 NSFG was conducted from September 2013 through September 2015, based on a survey protocol and informed consent procedures approved by the NCHS Research Ethics Review Board (protocol #2011-11). After sample respondents were selected based on screening interviews in selected households, in-person interviews were conducted with 5,699 women and 4,506 men 15-44 years of age for a total sample size of 10,205. The interviews were conducted by female interviewers trained specifically for the NSFG survey using laptop computers--a procedure called computer-assisted personal interviewing (CAPI). Informed consent related materials for the 2013-2015 NSFG are available on the NSFG website. Parental permission and minor assent were obtained for all minor respondents aged 15-17. The interviews for female respondents averaged 74 minutes in length, and the interviews for male respondents averaged 51 minutes, both within the limits of 80 minutes for females and 60 minutes for males approved by the Office of Management and Budget (NSFG OMB No. 0920-0314), The overall response rate was 69.3%; 71.2% for women and 67.1% for men. The response rate for male teenagers ages 15-19 was 71.2% and 70.7% for female teenagers. The calculation of response rates for the NSFG under the continuous design, beginning with the 2006-2010 data, has been similar for each release. Specific information for the 2006-2010 release can be found in http://www.cdc.gov/nchs/data/series/sr_02/sr02_158.pdf; specific information for the 2011-2013 release can be found in the documents linked in the next paragraph. Response rates for 2013-2015 were calculated in the same way as described in those

documents. Information on how response rates were calculated for earlier data files, including those for 1988 and 1995 when the NSFG was a subsample of women whose households participated in the National Health Interview Survey (NHIS), is available in the Series 2 reports for each file release available on the NSFG webpage here
http://www.cdc.gov/nchs/nsfg/nsfg_products.htm

Additional details on how the survey was designed and conducted are forthcoming at the time of this writing. However the principles of continuous interviewing and many ongoing design features since it began in 2006 can be found on the NSFG website including the methodology reports for the 2011-2013 fieldwork period here:

- **Summary of 2011-2013 Design and Data collection**

- **Sample design documentation for 2011-2013**

- **Sample error estimation for 2011-2013**

- **Weighting design documentation for 2011-2013**

WHERE TO FIND NSFG PUBLIC USE DATA AND DOCUMENTATION and OTHER NSFG DATA FILES

The public use data and documentation for the 2013-2015 NSFG are available on the NSFG website (www.cdc.gov/nchs/nsfg.htm). Documentation includes this User's Guide, the questionnaires, and a link to the interactive, online codebook documentation ("Webdoc") which shows separate entries for each variable on the files.

The public use data are contained in three data files:
- **Female respondent** file (one record or observation per interviewed female)
- **Female pregnancy (interval)** file (one record per pregnancy of interviewed females)
- **Male respondent** file (one record per interviewed male)

As was done for the 2011-2013 NSFG:

- The Audio Computer-Assisted Self-Interviewing (ACASI) data for 2013-2015 are included as part of the main public use files.

- The four-category region of residence variable (REGION) is available upon request, using a special data user agreement (see NSFG webpage for details).

- An NSFG-respondent-based file with the state and county FIPS codes at time of interview and state FIPS code at the time the last Census is accessible through the NCHS Research Data Center to enable researchers to merge the NSFG survey data with other contextual data.

A file of interviewer observations is forthcoming at the time of the file release.

ORGANIZATION OF THE 2013-2015 NSFG PUBLIC USE DATA FILES

The public use data for the 2013-2015 NSFG are provided as 3 separate ASCII files. The file characteristics of these ASCII files are shown below.

FILE CHARACTERISTICS	Number of Records (observations)	Record Length (number of columns)	Number of Variables
Female respondent file File = 2013_2015_FemRespData.dat (one record per woman)	5,699	5,084	3,207
Female pregnancy (interval) file File = 2013_2015_FemPregData.dat (one record per pregnancy reported by female respondents)	9,358	477	278
Male respondent file File = 2013_2015_MaleData.dat (one record per man)	4,506	4,480	2,940

The **Female Respondent file** contains one record for each of the 5,699 women interviewed in the survey and includes most of the information from their interviews.

The **Female Pregnancy (Interval) file** contains one record for each of the 9,358 pregnancies reported by female respondents. Pregnancy records are based on both completed pregnancies (those that have reached an outcome such as live birth, stillbirth, ectopic, miscarriage, or induced abortion) and current pregnancies (ongoing at time of interview). Each pregnancy record contains information about the characteristics of that pregnancy and method use and wantedness before that pregnancy. That is, in the Female Respondent file the unit of analysis is the woman, and in the Pregnancy file the unit of analysis is the pregnancy or pregnancy interval.

The **Male Respondent** file, often just referred to as the male file, contains one record for each of the 4,506 men interviewed; in this file the male respondent is the unit of analysis.

These ASCII data files can be read into any statistical software system for analysis purposes. Program statements to read the data into SAS, SPSS, and Stata, including variable and value labels, are provided on the NSFG webpage. Formats are provided within the program statements, but are for user convenience or ease of display only. They do not always reflect the actual values in the dataset and sometimes condense responses into categories. Data users should make their own decisions about whether to use the formats provided in the program statements. In addition, SAS and Stata syntax guidelines are provided in **Appendix 2** for common file manipulations such as combining data for males and females or combining female respondent-level and pregnancy-level data.

Data Layout for Each File

The following is a listing of the major sections and key variables contained in the three public use files from the 2013-2015 NSFG. For more detail on data file contents than what is shown below, see **Appendix 1** (File Indexes) in which short descriptions and variable types are provided for every variable included on the public use files.

FEMALE RESPONDENT FILE – *information for each female respondent interviewed*

- Respondent ID (CASEID) and selected screener variables
- Questionnaire Data (including computed variables) for Sections A-J
    - A: Background and demographic information
    - B: Pregnancy and adoption-related information
    - C: Marital and relationship history; first sexual intercourse; recent partners
    - D: Sterilizing operations and impaired fecundity
    - E: Contraceptive history and related information
    - F: Family planning and medical services
    - G: Birth desires and intentions
    - H: Infertility services and reproductive health
    - I: More background, more demographic information, access to health care & attitude questions
    - J: Audio CASI: general health measures; pregnancy reporting; cigarette, alcohol & other drug use; STD/HIV-risk behaviors; non-voluntary intercourse; sexual orientation & attraction; income and economic insecurity
- Recodes (constructed, imputed variables) & imputation flags (including key recodes describing pregnancies)
- Weights & related variables *(including sample design variables)*
- Date of interview and related variables

FEMALE PREGNANCY (INTERVAL) FILE – *information for each pregnancy reported by female respondents (including current)*

- Respondent ID (CASEID)
- Pregnancy Order (PREGORDR)

- Questionnaire Data (including computed variables) for Sections B & E
  - B: pregnancy outcomes and dates, prenatal care, sources of payment for delivery, breast-feeding
  - E: contraceptive use in the pregnancy interval and wantedness of the pregnancy
- Recodes (constructed, imputed variables) & imputation flag (including key recodes and other variables from respondent file)
- Weights & related variables *(including sample design variables)*
- Date of interview and related variables

MALE RESPONDENT FILE – Information for each male respondent

- Respondent ID (CASEID) and selected screener variables
- Questionnaire Data (including computed variables) for Sections A-K
  - A: Background and demographic information
  - B: Sex education, vasectomy & infertility, sexual intercourse, enumeration and relationship with up to 3 recent (or last) sexual partners
  - C: Current wife or cohabiting partner: date of marriage; children; contraception with her
  - D: Recent (or last) sexual partner(s) (up to three): key dates, children; contraception with her; 1st partner
  - E: Former wives and first premarital cohabiting partner: key dates, children; contraception with each
  - F: Other biological and adopted children; other pregnancies
  - G: Fathering: Activities with the youngest child he lives with and the youngest child he lives apart from
  - H: Desires and intentions for future children
  - I: Health conditions, access to health care, and receipt of health services
  - J: More background, more demographic information, & attitude questions
  - K: Audio CASI: pregnancy reporting; cigarette, alcohol & other drug use; STD/HIV-risk behaviors; non-voluntary intercourse; sexual orientation & attraction; income and economic insecurity
- Recodes (constructed, imputed variables) & imputation flags
- Weights & related variables *(including sample design variables)*
- Date of interview and related variables

SAMPLE WEIGHTS AND VARIANCE ESTIMATION

Since the NSFG data are based on a multi-stage probability-based, nationally representative sample of the household population aged 15-44, and not a simple random sample of the population, data users should understand how to account for the complex sample design when doing their analyses in order to obtain statistically valid results. This section provides a summary of the procedures used for sample weighting and variance estimation. Given that the sample design for the 2013-2015 NSFG is largely the same as that for the 2006-2010 and 2011-2013 files, much of the information published for those file releases is applicable for 2013-2015.

At the time of this writing, updates of this information are being prepared for the 2013-2015 NSFG, but the 2011-2013 documentation referenced earlier, on page 10, provides relevant information.

Sampling Weights

In the 2013-2015 NSFG, each person in the NSFG sample represents a different number of people in the U.S. household population and this number is indicated in the respondent's sampling weight. There are several factors that lead to variation in the size of the weights. For example, Hispanics, blacks, and teens were selected at higher rates than others in the 15-44 age group. Women also had a higher probability of selection than men. Sampling weights adjust for these unequal probabilities of selection for different population subgroups. The sampling weights were further adjusted to account for differential response rates and coverage rates, so that accurate national estimates can be made from the sample. The weights were adjusted to U.S. Census Bureau projections of the number of persons in age-sex-race-ethnicity subgroups. Data users should use the weights in all analyses to obtain accurate estimates. Using the weights will permit replication of the nationally representative estimates that appear in published NCHS reports.

Each of the 2013-2015 data files have a weight variable called "WGT2013_2015" with values for each of the 5,699 female and 4,506 male respondents who completed NSFG interviews in 2013-2015 . When correctly applied for the full set of cases, this "WGT2013_2015" variable yields estimates representative of the 61 million men and 61 million women in the household population aged 15-44 of the United States in each dataset at the approximate midpoint of 2013-2015 interviewing (July 2014). The weighted population sizes will of course be smaller if the user conducts analyses for other population subgroups.

As described in **Appendix 2**, there is also a weight variable "WGT2011_2015" released in a separate file that should be used when analyzing the combined 2011-2013 and 2013-2015 files to produce estimates for that four-year period. This sample weight is adjusted to the U.S. Census Bureau population projections at the approximate midpoint of the 2011-2015 time period, July 2013. Users who want to compare estimates for the 2011-2013 file to those from the 2013-2015 file would need to use the separate two-year weights for each file.

To yield the population number in thousands, as often appears in NCHS reports, you divide the sample weight by 1,000. For example, you would use this data manipulation step to create a new weight variable:

WGT1000= WGT2013_2015/1000

In addition to using the sampling weight variable WGT2013_2015, researchers must use the design variables for the sampling stratum (SEST) and cluster (SECU) to obtain correct standard errors for their estimates. These values of SEST and SECU would be used regardless of whether the user is analyzing the 2 year files (with the 2 year file weights) or the combined 4 year file (with the 4 year case weights).

Variance Estimation

Sampling variance is a measure of the precision of a statistic (such as a percentage, proportion or a mean) due to having taken a sample of instead of interviewing the full population. In the 2013-2015 NSFG, the sampling variance measures variation caused by interviewing the NSFG sample of 10,205 respondents instead of all 122 million women and men aged 15-44 in the US household population.

Many statistical software packages by default, compute "population" variances, which may severely underestimate or overestimate the sampling variances because they assume that the sample was drawn using simple random sampling. Statistical software that can analyze data drawn from complex survey samples is required to accurately estimate sampling errors in a complex sample such as the NSFG. For example, SAS procedures such as "FREQ" produce population variances assuming simple random sampling, but SAS has procedures for complex survey estimates in its 'SURVEY' procedures such as "SURVEYFREQ." Similarly, SUDAAN, Stata and SPSS have procedures designed to analyze data derived from a complex sample survey.

Selected examples of how to use such statistical software to estimate sampling errors based on 2013-2015 NSFG data are provided in **Appendix 2**. Three variance estimation examples using SAS and Stata for the 2013-2015 file are also posted on the NSFG webpage.

When estimating variances for population subgroups (such as those who have ever had sexual intercourse or those 30-44 years of age), it is important to read in the entire data set first. An indicator variable for your subpopulation (e.g. in SAS for 30-44 year olds, "if ager >= 30 then agenew=1") should be created to identify only those observations that will be used in the analysis. Then, you define your subgroup of interest within your survey procedure, such as using the SUBPOPN command in SUDAAN, the SUBPOP command in STATA, or by using the DOMAIN statement in SAS or including the variable in your SURVERYFREQ table command. If the data are subset without first reading in the entire data set, then empty clusters may be lost, making the survey design structure incomplete. This may result in the statistical software to terminate or lead to other errors.

OVERVIEW OF DATA QUALITY IN THE NSFG

As measured by amounts of missing data and inconsistent data, data quality in the 2013-2015 NSFG is high, as it was in previous data years. This high quality was obtained through:

-- Questionnaire design work, including careful specification, testing, and incorporation of lessons learned from the past NSFG data collection periods;
-- Consistency checks built into the interview that allowed potential data problems to be resolved in the field rather than after data collection;
-- Evaluation of monthly data files to find and correct instrument problems before significant numbers of cases were affected; and
-- Extensive interviewer training to ensure adherence to consistent and ethical fieldwork procedures.

Part 2 of the User's Guide provides information on those few variables where data issues exist.

Data files as large and complex as these cannot be guaranteed to be free of errors. If you believe you have found an error or need further assistance, please email the NSFG staff at NCHS at nsfg@cdc.gov.

DATA PREPARATION FOR PUBLIC USE

This section describes steps taken to prepare the NSFG interview data for public use. Some of these actions were taken simply to make the data more useful. Other actions were taken to protect the confidentiality of individual respondents, in keeping with the legal obligation of NCHS when conducting the NSFG or any of its other surveys.

Logical Inconsistencies and Out-of-Range Values

During fieldwork, logical consistency across data items was maintained through "edit checks" built into the programs that ran the male and female questionnaires. These edit checks alerted the interviewer to inconsistent or out-of-range entries and required that she attempt to correct the entry, usually by working with the respondent. Out-of-range values are greatly reduced in the 2013-2015 NSFG (as in past files) because valid ranges are specified and programmed into the instrument to the extent possible, and values outside that range are rejected or signaled by the computer.

Some edit checks in the instrument are **"hard edits"** in that they disallow combinations of values that are impossible (for example, respondents cannot report a date for any event in their lives that is later than the interview date or before their date of birth). Other edit checks are **"soft edits"** in that they alert the interviewer to situations that are rare but not impossible (for example, a respondent reports that she had her first menstrual period at a particularly young age).

In soft edit checks, the respondent is given the opportunity to revise his or her responses in case they were given in error. If the respondent says that the information is accurate, the interviewer is allowed to override or suppress the inconsistency warning box and enter a brief comment to explain the situation. In rare cases, the interviewer herself may have misunderstood the edit check and mistakenly suppressed it. In all such cases, the seemingly inconsistent data may remain on the data file. It is not possible to foresee and specify all the edit checks that might be needed in these very complex interviews, and as a result, some inconsistencies in the data could not be eliminated.

In addition to edit checks, other aspects of the questionnaire designed to maximize consistency *during* data collection were: 1) "summary screens" before or after key sections, reminding the respondent of events and dates reported earlier, and 2) life-history calendars provided to female respondents as a visual aid for recording and remembering the chronology of events.

In the 2013-2015 NSFG, the process of checking for consistency was focused primarily

on the recoded variables and variables used to construct them. These were considered to be the most critical and most frequently used variables in the files. Considerable efforts were made to detect and resolve or document inconsistencies and unacceptable codes throughout the files. However, as noted earlier, given the size and complexity of these data files, they may not be free of inconsistent or missing responses.

Coding for "Don't Know," "Refused," and "Not Ascertained" Values

Missing data refers to responses of "don't know" or "refused" that were entered by the interviewer to indicate that the respondent could not or would not provide an answer to a question. "Not ascertained" refers to rare instances in which a question was erroneously skipped during the interview. The code for "not ascertained" was generally assigned in these cases after fieldwork was completed. Only completed cases are in the files; a case was defined as being complete if the respondent answered the last applicable question before ACASI (in Section I for females and in Section J for males). The small number of respondents who did not complete the ACASI section, partially or completely, will have "not ascertained" values assigned to all variables after their break-off point.

Depending on the column length of the original data items:

- "don't know" values are coded 9, 99, 999, 9999, or 99999
- "refusal" values are coded 8, 98, 998, 9998, or 99998
- "not ascertained" values are coded 7, 97, 997, 9997, or 99997

(The codebooks only show these codes for the variable if any cases had those particular values.)

Missing data as described above is distinct from a variable that was inapplicable -- the respondent was legitimately skipped past the question (for example, respondents who had never been pregnant were not asked questions about how their pregnancies ended). For more information on determining who was asked each question, refer to the description of universe statements in the User's Guide section entitled "**Description of Codebooks**" further below or the codebook entry for particular variables. A question that was legitimately skipped or a variable legitimately not defined for a respondent will be coded as blank, and in the codebook, is indicated by a "dot" and labeled "inapplicable" or "sysmis."

Recoded variables may have legitimate inapplicable values, but they have no missing data in the form of "don't know," "refused," or "not ascertained" values because these responses were imputed to a valid value. Cases that had recode values imputed because of missing information on the source variables are identified with an imputation "flag"-- a separate variable that indicates whether or not the corresponding recode was imputed (see User's Guide section on "**Recodes and Imputation**" further below, as well as **Appendix 3** (Recode Specifications).

Century Month Coding for Dates

During the interview, dates of events were collected as month and year, except for the

respondent's date of birth, which was collected as month, day, and year. For every date asked in the interview, including the respondent's date of birth, the month and year information was converted to "century months" by <u>subtracting 1900 from the year, then multiplying the remainder by 12, and adding the number of the month</u>, where January = 1, February = 2, and so on. *(Note that the respondent's <u>day</u> of birth is not available on the public use file. His or her birthday is converted to the century month variable cmbirth.)*

For instance:
The century month code for February 1959 is (59 x 12) + 2= 710.

The century month code for October 1987 is (87 x 12) + 10 = 1054.

The century month code for January 2000 is (100 x 12) + 1 = 1201.

The century month code for July 2006 is (106 x 12) + 7 = 1279.

The century month form is convenient for computing intervals between dates, and subtraction yields intervals in months.

With the exception of one recoded date variable (DATEUSE1 on the female respondent file) that has a leading 9 to indicate when the value was estimated, all century month date variables in the file are 4 columns long. The following codes were used for the 3 types of missing data on century month date variables:

9997 = Not ascertained
9998 = Refused
9999 = Don't know

When seasons are reported on any month variable in the NSFG interview, a particular month is consistently assigned to enable the construction of a century month value and facilitate subsequent routing through the questionnaire. Listed below are the months assigned when a season is reported:

Winter = 1 (January)
Spring = 4 (April)
Summer = 7 (July)
Fall   = 10 (October)

If a respondent says "don't know" or refuses (DK/RF) to report a month, the value "6" (June) is assigned for the month. If a respondent does not report a year, the century month variable is set to 9999 for "Don't Know" or 9998 for "Refused." Data users may wish to refer to the raw month and year variables included on the public use files, particularly the month variable, if they encounter an implausible or less likely set of century month dates.

The century month codes from 361 (January 1930) through 1392 (December 2015) are shown in the array below with the years from 1930 through 2015 on the vertical axis and the months on the horizontal axis. The code for a given month and year can be found by reading

across the line for the appropriate year to the column headed by the appropriate month.

All interviews for the 2013-2015 NSFG were conducted between September 2013 (century month 1365) and September 2015 (century month 1389).

CENTURY MONTH CODES

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1930	361	362	363	364	365	366	367	368	369	370	371	372
1931	373	374	375	376	377	378	379	380	381	382	383	384
1932	385	386	387	388	389	390	391	392	393	394	395	396
1933	397	398	399	400	401	402	403	404	405	406	407	408
1934	409	410	411	412	413	414	415	416	417	418	419	420
1935	421	422	423	424	425	426	427	428	429	430	431	432
1936	433	434	435	436	437	438	439	440	441	442	443	444
1937	445	446	447	448	449	450	451	452	453	454	455	456
1938	457	458	459	460	461	462	463	464	465	466	467	468
1939	469	470	471	472	473	474	475	476	477	478	479	480
1940	481	482	483	484	485	486	487	488	489	490	491	492
1941	493	494	495	496	497	498	499	500	501	502	503	504
1942	505	506	507	508	509	510	511	512	513	514	515	516
1943	517	518	519	520	521	522	523	524	525	526	527	528
1944	529	530	531	532	533	534	535	536	537	538	539	540
1945	541	542	543	544	545	546	547	548	549	550	551	552
1946	553	554	555	556	557	558	559	560	561	562	563	564
1947	565	566	567	568	569	570	571	572	573	574	575	576
1948	577	578	579	580	581	582	583	584	585	586	587	588
1949	589	590	591	592	593	594	595	596	597	598	599	600
1950	601	602	603	604	605	606	607	608	609	610	611	612
1951	613	614	615	616	617	618	619	620	621	622	623	624
1952	625	626	627	628	629	630	631	632	633	634	635	636
1953	637	638	639	640	641	642	643	644	645	646	647	648
1954	649	650	651	652	653	654	655	656	657	658	659	660
1955	661	662	663	664	665	666	667	668	669	670	671	672
1956	673	674	675	676	677	678	679	680	681	682	683	684
1957	685	686	687	688	689	690	691	692	693	694	695	696
1958	697	698	699	700	701	702	703	704	705	706	707	708
1959	709	710	711	712	713	714	715	716	717	718	719	720
1960	721	722	723	724	725	726	727	728	729	730	731	732
1961	733	734	735	736	737	738	739	740	741	742	743	744
1962	745	746	747	748	749	750	751	752	753	754	755	756
1963	757	758	759	760	761	762	763	764	765	766	767	768
1964	769	770	771	772	773	774	775	776	777	778	779	780
1965	781	782	783	784	785	786	787	788	789	790	791	792

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1966	793	794	795	796	797	798	799	800	801	802	803	804
1967	805	806	807	808	809	810	811	812	813	814	815	816
1968	817	818	819	820	821	822	823	824	825	826	827	828
1969	829	830	831	832	833	834	835	836	837	838	839	840
1970	841	842	843	844	845	846	847	848	849	850	851	852
1971	853	854	855	856	857	858	859	860	861	862	863	864
1972	865	866	867	868	869	870	871	872	873	874	875	876
1973	877	878	879	880	881	882	883	884	885	886	887	888
1974	889	890	891	892	893	894	895	896	897	898	899	900
1975	901	902	903	904	905	906	907	908	909	910	911	912
1976	913	914	915	916	917	918	919	920	921	922	923	924
1977	925	926	927	928	929	930	931	932	933	934	935	936
1978	937	938	939	940	941	942	943	944	945	946	947	948
1979	949	950	951	952	953	954	955	956	957	958	959	960
1980	961	962	963	964	965	966	967	968	969	970	971	972
1981	973	974	975	976	977	978	979	980	981	982	983	984
1982	985	986	987	988	989	990	991	992	993	994	995	996
1983	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008
1984	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020
1985	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032
1986	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044
1987	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056
1988	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068
1989	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080
1990	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092
1991	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104
1992	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116
1993	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128
1994	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140
1995	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152
1996	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164
1997	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176
1998	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188
1999	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200
2000	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212
2001	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224
2002	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236
2003	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248
2004	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260
2005	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272
2006	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284
2007	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296
2008	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308
2009	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320
2010	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332
2011	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344
2012	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356
2013	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368
2014	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380
2015	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392

Other-Specify Coding

In the 2013-2015 NSFG, as in past surveys, a small number of questions contained items to which respondents could specify a response other than those provided, and this response was typed in verbatim by the interviewer. In most cases, these questions appear in the questionnaires with a "_SP" at the end of their variable name or a "SP_" at the beginning, and the question text and response type distinguish these as open-ended questions. These verbatim response variables are NOT included on the public use files because of confidentiality concerns, but the essential information for data users has been coded in a manner that does not pose risk of disclosure. In all cases, responses that were clearly codable using a pre-specified category (one of the categories offered in the questionnaire and shown to respondents on a show card) were edited (or "back-coded") into that pre-existing category. In a few instances, the responses were categorized, and a new variable was created that contained only these additional responses.

In summary, the actual verbatim responses for other-specify variables are not included in the public-use data files, but they have all been reflected in some way in the associated, closed-ended numeric variables that are included on the public use file. Questions where new categories or new variables were created based on other-specify responses have a short description in a "Note" entry on their codebook page (see also **Description of Codebooks** further below). Please consult the questionnaires for further details on these questions with "other (specify)" options. Part 2 of the User's Guide also includes specific information related to some of these items.

Recodes and Imputation
*(also see **Appendix 1** (File Indexes), **Appendix 3** (Recode Specifications), and **Appendix 4** (Recode Crosswalks))*

In order to facilitate consistent, comparable estimates of key NSFG measures for all data users, NCHS produces a number of "recoded variables," or "recodes" for each public use file. Published NCHS reports use these recodes whenever available because they permit internally consistent and replicable estimates. NCHS also uses the recodes to prioritize the cleaning of the data file: there are too many variables in the data file to edit or reconcile them all, so NCHS focuses its cleaning and editing primarily on the recodes and on the variables that are used to construct the recodes. (Recodes comprise about 10% of the variables in these files.)

Some recodes are fairly simple, while others are quite complex. Some recodes may simply be transferred from single questionnaire items and imputed if missing (for example, RCURPREG, whether respondent is currently pregnant). Other recodes are based on multiple questionnaire items and may involve more intricate logic to define. (For example, CONSTAT1, the respondent's current contraceptive status)

Before using the original data items or constructing their own summary variables, analysts are encouraged to **check to see if a relevant recode exists**. Many of the raw or computed variables that have a recode corresponding to them will have a **note on their Webdoc codebook pages** stating the name of the appropriate recode.

For convenience, below is a list of some commonly used recodes corresponding to background characteristics and other key NSFG variables. Unless otherwise indicated, the recodes are available for males and females.

AGER	R's age at interview
FMARITAL	Formal (legal) marital status
RMARITAL	Informal marital status
EDUCAT	Education (number of years of schooling)
HIEDUC	Highest completed year of school or highest degree received
HISPANIC	Hispanic origin, regardless of race
RACE	Race of respondent, regardless of Hispanic origin
HISPRACE	Race and Hispanic origin – based on 1977 OMB guidelines
HISPRACE2	Race and Hispanic origin – based on 1997 OMB guidelines
INTCTFAM	Intact status of childhood family
PARAGE14	Parental living situation at age 14
EDUCMOM	Mother's (or mother-figure's) education
AGEMOMB1	Age of mother (or mother-figure) at first birth
METRO	Place of residence (metropolitan-nonmetropolitan)
RELIGION	Current religious affiliation
LABORFOR	Labor force status
POVERTY	Poverty level income
TOTINCR	Total income of R's family
PUBASSIS	Whether R received public assistance in the calendar year before the interview
HADSEX	Whether R has ever had sexual intercourse with opposite sex
VRY1STAG	Age at first intercourse
VRY1STSX	Date (century month) of first intercourse
CONSTAT1	Current contraceptive status (females only)

Besides the above list, other sources to check are the File Indexes in **Appendix 1** or the Recode Specifications in **Appendix 3** to see if a relevant recode exists. If you want to see whether there are comparable recodes between males and females or across NSFG data years, **Appendix 4** contains 3 crosswalks for this purpose:
- ■ 4a: recodes for males and females in 2013-2015 NSFG (arranged by topics)
- ■ 4b: recodes for females across 2002, 2006-2010, 2011-2013, and 2013-2015 NSFG (arranged by section of questionnaire)
- ■ 4c: recodes for males across 2002, 2006-2010, 2011-2013, and 2013-2015 NSFG (arranged by section of questionnaire)

The frequency of missing values for the recoded variables in 2013-2015 is quite low, just as it was in past files. Cases that have missing data on a recode (i.e., their values could not be constructed from the source variables referenced in the recode specifications) have been imputed.

Most missing recode values were assigned using **regression** imputation software in which multiple regression is used to predict a value for the case using other variables in the data set as predictors. Regression imputation follows the same logical constraints that are built into the original recode specifications. To the extent possible, imputed values are checked to ensure that the imputed values are within acceptable ranges, and are consistent with other recodes and other data reported by the respondent.

A smaller number of cases for some recodes were imputed using **logical** imputation, which involves NCHS staff examining variables related to the variable in question, and assigning a value that is consistent with those other variables.

Imputation flag variables were created for every recode, allowing users to determine whether the value for each case is based on reported data, or imputed data. They also indicate which kind of imputation was used. Each imputation flag has the following potential values:

    0=Questionnaire data (not imputed)
    1=Multiple regression imputation
    2=Logical imputation

A value of 0 on the imputation flag means that imputation was not necessary; the reported questionnaire data were sufficient to determine an appropriate value on the recode. All values other than 0 indicate that the case was imputed for this recode. Further details on the imputation process for 2013-2015 are forthcoming at the time of this writing, but since the process used for the 2013-2015 release is similar to that used in the prior release, the below report (for the 2011-2013 data) can be consulted for details:

- **Summary of 2011-2013 Design and Data collection**

As noted above, all recodes have been checked thoroughly against related data items and edited if necessary for consistency. Except when it was obviously incorrect, and involved critical or commonly used variable(s), actual reported information was never replaced by an imputed value. **NCHS recommends that analysts use all cases in the file, including those with recode values imputed**. Using weights and imputed cases will enable the analyst to replicate results that appear in NCHS reports. The impact of imputation on analyses can be examined by using the imputation flags to compare results with and without the imputed cases.

Finding recodes in the data file and codebook: As shown in **Appendix 1** (File Indexes), the recodes and their imputation flags are clustered together near the end of each of the 3 data files. Recodes can be distinguished in the codebook documentation by the "Variable type" displayed on the codebook page between the variable name and the variable description. The word "recode" also appears at the end of the variable's "question text" or short description.

Protections to Minimize Risk of Disclosure for Individual-Level Data

When NCHS collected data from respondents for the NSFG, those respondents were promised in the informed consent process that the information they provided would be kept

confidential. NCHS is legally required to keep that promise, both during fieldwork and in the production of data files for public use.

Prior to their release, the proposed NSFG public use files were reviewed by the NCHS Disclosure Review Board (DRB) and the NCHS Confidentiality Officer. The NSFG staff included a number of modifications to the data files to prevent disclosure of the identities of the respondents, and at the same time preserve the analytical value of the data.   In response to the DRB's review, the NSFG staff and contractor made some further changes to minimize the risk of disclosure, including the suppression or collapsing of additional variables that could be used to identify very small groups.   In brief:

- All directly identifying information (such as names and addresses) has been eliminated from the public use files.

- The only geographic variable included on the public use files is a 3-category METRO recode (principal city of Metropolitan Statistical Area (MSA), other MSA, not MSA).

- All variables on the files that could potentially be used to indirectly identify individuals have been recoded or re-categorized, with particular attention to keeping categories that are substantively useful, and collapsing categories that were so small that they were of limited analytical use.   When variables have been recoded or (or modified in some other way) for this purpose, a note is included at the end of the universe statement in the online codebook, Webdoc (described in the next section). For example, the variable, AC-2 HISPGRP, shows the following universe statement: "Universe : Applicable if R is Hispanic (AC-1 HISP = 1) (This variable has been recoded for public use.)"

As a final step to prevent identification of individual respondents, the values of some variables have been altered for some respondents through **statistical perturbation**. That is, a few values in the data set are no longer the actual values reported by the respondents, resulting in greater uncertainty for anyone attempting to identify a particular individual they may know participated in the survey. However, these alterations, or statistical perturbations, were carefully designed to give analysts comparable statistical information as those obtained from the unaltered responses. In other words, it is extremely unlikely that either national estimates or causal models are affected by any of the alterations, except for a very slight increase in the variance of a few statistics.

For information on the issues and techniques related to disclosure limitation and confidentiality, please consult literature such as the following:

Doyle P, Lane J, Theeuwes JJM, and Zayatz LV, editors. 2001. Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies. New York: Elsevier.

Federal Committee on Statistical Methodology. (May 2005). *Report on Statistical Disclosure Limitation Methodology*. (Statistical Policy Working Paper 22). Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. http://www.fcsm.gov/working-papers/spwp22.html

Muralidhar K, and Sarathy R. 2003. A Theoretical Basis for Perturbation Methods.   Statistics and Computing 13(4):329-35.

The individual-level variables that could not be included on the public use NSFG files, or could not be included in their original form, are available to the research community through the NCHS Research Data Center (RDC). The full list of these variables is provided in **Appendices 7a and 7b**, and further details on the specific modifications made to these variables are provided in **Appendix 7c**. As with all data files available through the NCHS RDC, these restricted-use data are made available to researchers under special arrangements that assure confidentiality and protection of the data.

Researchers who wish to learn more about or apply for access to any of these NSFG files available through the Research Data Center – Contextual Data, Restricted-Use individual-level data, or Interviewer Observation data -- should first look at information provided on the RDC website (www.cdc.gov/rdc), and then contact either the NSFG staff at nsfg@cdc.gov or the RDC at rdca@cdc.gov.
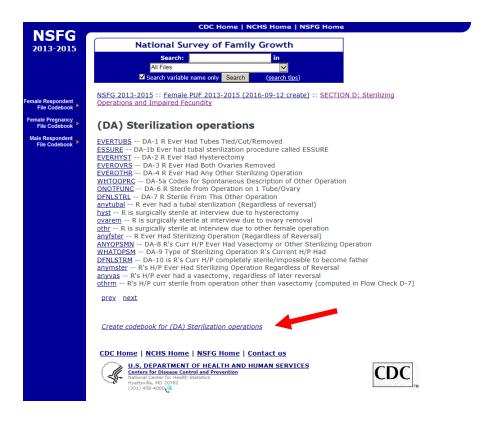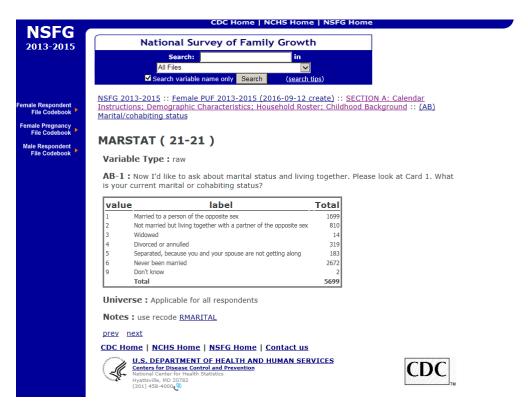
DESCRIPTION OF CODEBOOKS

Overview

Codebooks for the NSFG provide essential information for each variable included in the public use files. The elements of the codebook, such as variable type and universe statements, are described further below.

Webdoc is a web-based tool that permits online, interactive access to the NSFG codebooks and allows easy access to all variables, quick navigation between different sections of the instrument (through hyperlinks and through menu-like lists of data files, sections and series), and searching for specific words or phrases or for specific variables. For recoded variables, direct links are provided to the recode specifications. These recode specifications are the same as those provided in **Appendices 3a-3c** on the NSFG webpage.

To generate an excerpt for specific sections or subsections of the codebook use the hyperlink at the bottom of the screen that says "Create codebook for <Section/Subsection>." The resulting file can be saved in PDF, printed, or simply viewed on the screen. For example, at the bottom of the Webdoc page listing the variables in sub-section, "Sterilization Operations (DA)" if you click on *Create codebook for (DA) Sterilization Operations,* a screen with just the variables in the DA series from Female Section D is created.

Below is an example page from Webdoc displaying the detailed codebook information for the raw variable (and question) AB-1 MARSTAT. The specific elements of the codebook are described further below.

Elements of the Codebook

Each variable in the public use files is represented in the codebook documentation with a page or entry containing all of these elements, which are described in turn below:
- Variable name
- Variable type
- Column locations
- Question text
- Universe statement
- Response categories and unweighted frequencies
- Link to recode specifications, where applicable
- Notes, where applicable

Variable Name:

For raw and computed variables, the variable name corresponds in most cases exactly to the question or computed variable name that appears in the CAPI Reference Questionnaire (CRQ). *(For example, in the Webdoc screen capture above, the variable name and question name are the same – MARSTAT.)* Recode and intermediate variable names correspond to those found in the recode specifications (**Appendix 3**). Throughout the codebook documentation and in the recode specifications, raw and recode variables are in uppercase and computed variables are in lowercase. In some cases where questions or variables are applicable for "loops" or "arrays" (such as pregnancies, marriages, months of the year, mentions for "enter all that apply" questions, etc.), then the variable names seen in the CRQ will have numeric suffixes attached. For example, BD-8 PAYBIRTH in the CRQ for Female Section B is the question asking how the delivery costs were paid for this child's birth, and respondents could select all options that applied. In the 2013-2015 pregnancy file, women reported no more than 3 forms of payment for their deliveries although space was allowed for 4 mentions, so the file includes 3 variables for those 3 mentions in PAYBIRTH1-3, and this information is noted in the variable labels and question text also.

Variable Type:

On the codebook page for each variable, underneath the variable name, is "Variable Type." There are four basic variable types included in the NSFG files -- "raw," "computed," "recode," or "intermediate":
- A **raw** variable refers to a question that was asked during the interview (the majority of variables in the data files are raw). *(For example, in the Webdoc screen capture above, AB-1 MARSTAT is labeled as a raw variable.)*
- A **computed** variable is a variable that is either computed as part of the Blaise-programmed survey instrument or constructed in post-processing after data collection, from one or more raw variables. Computed variables do not have missing values imputed.

- A **<u>recode</u>** variable is a constructed variable created after the data are collected, from one or more raw variables, and has missing values imputed.
- An **<u>intermediate</u>** variable is one defined in the specifications for certain recodes. These are sometimes included in the public use data files because they can be useful for analysts. It differs from a "computed" variable because it is neither constructed <u>during</u> the interview, nor is it defined solely as a post-processing variable. Intermediate variables are defined in the process of constructing a recode. They are very few in number.

Column Locations:

Next to each variable name, the codebook page gives the column locations in parentheses. *(For example, in the Webdoc screen capture above, the column location for MARSTAT is 21 on the female respondent file.)* These are primarily relevant if you wish to read in only selected variables from the ASCII data files. Program statements for SAS, Stata, and SPSS are provided on the NSFG webpage to read in each of the three NSFG data files for 2013-2015, including statements to assign appropriate variable and value labels.

Question Text:

Question text is either the actual question wording for a raw, <u>asked</u> question in the interview or the description of a computed, recode, or intermediate variable. The wording of the survey question is shown in the codebook for "raw" variables and is preceded by the question number. Any question wording variants are presented, sometimes in collapsed form. For computed variables (computed as part of the programmed survey instrument or in post-processing), the "question text" is the same as the variable's short description seen in the **File Indexes (Appendix 1)**, except that the "Flow Check" number has been included from the CAPI Reference Questionnaire (CRQ) where the computed variable was defined. For example, "(Computed in Flow Check E-13b)" indicates that the variable is defined in Section E, Flow Check E-13b. CRQs can be consulted to see these flow checks and the specifications for defining each computed variable. For recodes and intermediate variables (variables not represented in the questionnaires), the question text corresponds to the variable's short description from the Recode Specifications (see **Appendix 3**).

When variables are part of an array or loop, the question text on the codebook page indicates what is being referenced, just as the variable descriptions do in the File Indexes (**Appendix 1**). For example, the question text for male CG-5 CWPCHSEX2 makes clear that the variable applies to the second biological child the respondent had with his current wife or cohabiting partner.

Universe Statements ("Applicable Specifications"):

In the codebook documentation, the "applicable specifications" or "universe statement" for a question indicates which respondents were asked the question or had the variable defined

for them. If a question or variable was not applicable to a particular respondent, the questionnaire program skipped to the next applicable question. If a question was not skipped by any respondent or the variable was assigned a non-blank value for every case, the universe statement says, "Applicable for all respondents" or for the pregnancy file, "Applicable for all pregnancies." *(For example, see Webdoc screen capture for AB-1 MARSTAT above.)* One other type of information that is appended to the end of some universe statements is whether the variable has been recoded (or modified in some other way) for public use. This relates to modifications that were made to some variables to minimize the risk of disclosure (see section on "Protections to Minimize Risk of Disclosure for Individual-Level Data"), and these notes will only appear if applicable. For example, AC-2 HISPGRP shows the following universe statement:

> Universe : Applicable if R is Hispanic (AC-1 HISP = 1) (This variable has been recoded for public use.)

Inapplicable variables are coded as blanks:

Some computer programs such as SAS and Stata read a blank as a non-numeric character (a dot) or "system missing" value, but others may read it as a zero. Analysts using statistical packages other than SAS or Stata should take care to distinguish between missing values and zeroes in programs used with these data because zeroes are often valid values on NSFG variables.

For many variables in the NSFG files, an *abridged* version of the complete universe statement is provided with the core routing information. These variables have nested routing statements, and for these variables, the most proximate routing statement will be described in the universe statement. Since the universe statement contains the variable(s) that determined the routing into that question, users can trace back through the routing logic, that is, go to each preceding variable to see its routing statement and continue until the universe statement reads, "Applicable for all respondents."

For example: the question EA-12 ECTIMESX in the female questionnaire reads, "How many different times have you used emergency contraception?" It was asked of those who had ever used emergency contraception. Thus, EA-11 MORNPILL ("whether R ever used emergency contraception") is included in the universe statement for EA-12 ECTIMESX, and hyperlinked. Clicking on this variable takes you to the codebook page for EA-11 MORNPILL. MORNPILL was only asked of those who had ever had sex, so its universe statement contains the computed variable for ever had sex: "rhadsex." Clicking on this takes you to the page for this computed variable which is "applicable for all respondents."

All public use file variables referenced in the universe statements are "hyper-linked" in Webdoc (see description above) so that users can go directly to their codebook pages. The variable names will not be hyperlinked in any PDF file generated from Webdoc, but users should still find it straightforward to find the relevant codebook entries for variables referenced in the universe statement. To make this easier to locate these variables, recall that question numbers precede the names of all raw variables. Also, the names of computed variables appear in lower case, and the names of raw variables and recodes appear in upper case. If a universe statement

references variables that are not included on the public use file those variables are not hyperlinked.

In addition to consulting the universe statement or "applicable specification" in the codebook documentation, you may also wish to consult:

- The CAPI Reference Questionnaire (CRQ; see section below entitled "**Description of Questionnaires**"), which contains more detailed specifications for the questionnaire. The universe statements for the computed variables are drawn from the Flow Checks in which the variables are defined. For the raw/asked variables as well as the computed variables, the questionnaires allow you to examine the sequencing and context of your variables of interest.

- The recode specifications (**Appendix 3**), which are the source for the universe statements included in the codebook, and also for the full details on how the recode was constructed and imputed.

Response Categories and Unweighted Frequencies:

For categorical variables and several continuous variables in the NSFG, the codebook documentation lists all values, as long as there are any cases with those values in the data files, with descriptive value labels and unweighted frequencies (or counts of cases). For example, if no one responded "don't know" to a particular item, the "don't know" value will not be displayed in the codebook To the extent possible, the exact wording of the questionnaire response choices is shown *(for example, see Webdoc screen capture for AB-1 MARSTAT above)*. Frequencies of variables that are not applicable for all respondents include the number of "inapplicable" cases. Most century month (date) and continuous variables have been collapsed *for display purposes* into more manageable groups, such as grouping individual century months into ranges of years. The original values of these variables are intact in the file unless otherwise indicated. As mentioned earlier, the program statements provided with the ASCII data files sometimes include statements to collapse values for display purposes and can be deleted or commented out. Data users should decide whether they want to use the format statements for specific variables. Response categories are not displayed unless at least 1 case reported such a response, and this also applies to "refused" (8, 98, etc.), "don't know" (9, 99, etc.), and "not ascertained" (7, 97, etc.) responses.

Link to recode specifications: For every recode variable, there is a direct link in Webdoc to the specification for that recode in **Appendix 3**. These hyperlinks will not be active in any generated PDF files for the codebooks.

Notes (where applicable): For selected variables, the codebook page will show a Note with further information. The primary reasons for these notes are to indicate when there is a relevant recode that you should use *(see note pointing to the RMARITAL recode in the Webdoc screen capture for AB-1 MARSTAT above)*, or to describe special circumstances related to the response categories (for example, see female HE-4 PLCHIV). Some notes also contain a

reference to this User's Guide, Part 2, if there is further information the user may find helpful when using the variable.

DESCRIPTION OF QUESTIONNAIRES

The NSFG webpage provides the male and female questionnaires in 2 formats:

-- CAPI-Lite format
-- CAPI Reference Questionnaire (CRQ) format

Both formats represent the basic content and routing of the full NSFG interviews, including the computer-assisted personal interviews (CAPI) administered by interviewers and the audio computer-assisted self-interviews (ACASI) that respondents completed on their own. However, each format of the questionnaire offers users a different level of detail on how the interview was conducted.

CAPI-Lite Format

The male and female interviews are shown in their entirety, but with abridged representations of the question wording variants and shorter descriptions of skip patterns through the interview. With this format, the emphasis is on getting a clear picture of how the questions were asked, in what order, and of which respondents, without showing every detail that was needed to program the questionnaire.

CAPI Reference Questionnaire (CRQ) Format

The CRQ shows all the detailed specifications that were used to program the NSFG questionnaires in Blaise Survey Software (see http://www.blaise.com/). While the entire female and male CAPI-Lites are contained in 2 PDF documents, the considerably longer CRQs are provided on the NSFG webpage as separate PDF files corresponding to each section of the questionnaire.

- All question wording variants are shown, along with the conditions defining when each variant should be used.

- "Flow Checks" specify the precise routing through the interview based on earlier questionnaire items so that the appropriate next questions for a particular respondent appear onscreen. In addition, in some instances flow checks include the creation of new variables from one or more of the "raw" or "asked" variables. These are called "computed variables" and are described in other sections of the User's Guide (see **Description of Codebook, "Variable Type"**). The flow check specifies in detail how these computed variables were defined. A summary list of computed variables defined in each questionnaire section can be found at the beginning of each section's CRQ, and those that are "passed forward" to be used for routing later in the interview are listed at the end of each section's CRQ.

- "Edit Checks," programmed into the instrument, attempt to catch and resolve data inconsistencies during the interview, rather than requiring resolution after data collection has ended. These consistency checks are generally located in the CRQ after the questions they are intended to reconcile. They are generally scripted for ease of use, and enable the interviewer to return to specific questionnaire items and correct them, if necessary. See also the User's Guide section on **Data Preparation for Public Use, "Logical Inconsistencies and Out-of-Range Values."**

- Use of additional survey aids, such as Show Cards, Help Screens, and the Life History Calendar (female interview only), is noted on individual questionnaire items. For example, if a question-specific help screen was available for an item, the CRQ indicates "[HELP AVAILABLE]." If the item's response choices were to be shown on a Show Card in the interviewer's show card booklet, the CRQ indicates the number of the show card along with the response categories. Also shown are the onscreen instructions for interviewers that accompanied many of the questions. An example of the Life History Calendar is shown on the next page. The version shown is for interviews conducted in 2015. Since the female interview includes a greater number of questions about dates and the relative timing of events than the male interview, the Life History Calendar was only used in the female interview. The Life History Calendar has been shown to improve recall of dates by anchoring responses to key life events for the respondent.

National Survey of Family Growth: Life History Calendar

Before 2012		2012												2013												2014												2015												
		Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc	Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc	Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc	Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc
	Your Age																																																
	Education																																																	
	Births & Other Pregnancies																																																	
	Marriages, Cohabs, Partners																																																	
	Intercourse																																																	
	Birth Control Methods																																																	
		Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc	Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc	Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc	Ja	Fe	Ma	Ap	My	Jn	Jl	Au	Sp	Oc	Nv	Dc

Your Date of Birth: _____

Date of First Intercourse: _____

Birth or Pregnancy Ending Dates:

1st: _____   2nd: _____   3rd: _____   4th: _____   5th: _____

USER SUPPORT

Most commonly asked questions about the NSFG are addressed in the "Frequently Asked Questions" included in this User's Guide in **Appendix 6** and throughout the main text of this User's Guide. If, however, you have reviewed this Guide and its appendices thoroughly and you still have a question, please contact the NSFG team:

By email: NSFG@cdc.gov
By phone: (301) 458-4222

ACKNOWLEDGMENTS

As noted in the survey background section above, the NSFG has been designed, administered, and disseminated by the National Center for Health Statistics since 1973, in collaboration with several other agencies of the U.S. Department of Health and Human Services (DHHS). (NCHS became part of CDC in 1987.) The 2013-2015 NSFG was jointly planned and funded by the following agencies within the DHHS:

- CDC's National Center for Health Statistics (NCHS)
- Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)
- Office of Population Affairs (OPA)
- Children's Bureau of the Administration for Children and Families (ACF/CB)
- Office of Planning, Research, and Evaluation within ACF (ACF/OPRE)
- CDC's Division of HIV/AIDS Prevention (DHAP)
- CDC's Division of STD Prevention (DSTDP)
- CDC's Division of Cancer Prevention and Control (DCPC)
- CDC's National Center for Birth Defects and Development Disabilities (NCBDDD)
- CDC's Division of Reproductive Health (DRH)

The NSFG team at NCHS holds primary responsibility for all aspects of the survey design, public use file and documentation preparation, and data dissemination as well as published reports and key statistics on the NSFG webpage. The NCHS NSFG team works closely with the contractor on the sample design and data collection for the survey. Fieldwork for the 2013-2015 NSFG was conducted under contract with the University of Michigan's Institute for Social Research (ISR) (Contract # 200-2010-33976) covering the survey data collection for 2011-2019.

An adequate acknowledgment of all those who made contributions at NCHS and at ISR to the design, conduct, and production of the 2013-2015 NSFG public use files would require several pages. This brief acknowledgement will only name those who made major contributions.

The NCHS NSFG team is comprised of Anjani Chandra (NSFG Team Lead and Principal Investigator), Joyce Abma (Contracting Officer Representative for contract with ISR), Gladys.

Martinez, Casey Copen, Kimberly Daniels, Isaedmarie Febo-Vazquez, Jill Daugherty, Chinagozi Ugwu, and Colleen Nugent. William D. Mosher and Jo Jones also contributed until their retirements in 2014 and 2015, respectively. Further consultation on other statistical matters at NCHS was provided by Van Parsons and Yulei He of the Division of Research and Methodology. During the survey period of the 2013-2015 NSFG, NCHS was under the direction of Charles J. Rothwell. The NSFG team at NCHS is housed within the Reproductive Statistics Branch (RSB) in the Division of Vital Statistics (DVS) at NCHS. Delton Atkinson is Division Director for DVS, and Amy Branum is Branch Chief for RSB.

Key NSFG contract personnel at University of Michigan's ISR, with their NSFG project roles in parentheses, include Mick Couper (Project Director), William G. Axinn (Deputy Project Director), Heidi Guyer (Field Director/Operations Manager), James Wagner (Senior Mathematical Statistician), Peter Granda (Director of Data Processing), and William Connett (Director of Information Security). Other significant contributors at ISR include Michael Shove, Stephanie Windisch, Karl Dinkelmann, Maureen O'Brien, and Patricia Berglund.

NCHS is grateful for the support and contributions of all those noted above, as well as the more than 100,000 NSFG survey respondents since 1973 who have given their time and energy to provide quality information for the survey.

Suggested Citation for NSFG Public Use Data and Documentation

As part of the federal statistical system, NCHS supports the dissemination of the NSFG public use data and documentation files at no charge to the public, and users in all research settings are encouraged to work with these data. NCHS appreciates citation when submitting research grant proposals or publishing their analyses because it will alert their audiences to the source of the data.

National Center for Health Statistics (NCHS). (2016). 2013-2015 National Survey of Family Growth Public Use Data and Documentation. Hyattsville, MD: CDC National Center for Health Statistics. Retrieved from http://www.cdc.gov/nchs/nsfg/nsfg_2013_2015_puf.htm

USER'S GUIDE PART 2:
TOPIC-SPECIFIC NOTES FOR ANALYSIS

In addition to these topic-specific notes, users can find information about questionnaire changes made between 2011-2013 and 2013-2015 in **Appendix 5**.

Abortion under-reporting among females

Abortions have always been under-reported in the NSFG and virtually all other demographic surveys. This has been determined by comparing NSFG weighted estimates of abortions with external data from abortion providers. The User's Guides of prior NSFG file releases show the percentages of underreporting based on these comparisons (see for example, the User's Guide for 2011-2013 NSFG). As in previous surveys, the NSFG staff advises NSFG data users that, generally speaking, NSFG data on abortion should **not** be used for substantive research focused on the determinants or consequences of abortion. The NSFG abortion data can be used for:

(1)   methodological studies of factors affecting abortion reporting.
(2)   studies of contraceptive efficacy, but only after the data are adjusted for the under-reporting of abortion.

Birth reporting among females

Data users interested in seeing how the births reported in the NSFG compare to those based on the vital records system should see the notes in part 2 of the 2011-2013 User's Guide for more information. An updated table with 2013-2015 NSFG data has not been produced at the time of the file release.

Religion data

See the 2011-2013 User's Guide notes on religion data in the NSFG, as no changes were made to these variables for 2013-2015 NSFG.

Use of Screener Data for HISPANIC, HISPRACE, and HISPRACE2

The recode variables HISPANIC, HISPRACE, and HISPRACE2 were used in the adjustment of the sample weights to population totals by Hispanic origin and race (along with age and sex). For the 2013-2015 files, these weight adjustments were done prior to the regression imputation of other recode variables so missing values needed to be assigned in another way. Information about Hispanic origin and race was logically imputed from screener information for these recodes.

FEMALE RESPONDENT FILE NOTES

Section C: Computed variables P1YRELPx

The purpose of these Blaise-computed variables is to capture the type of relationship with up to 3 sexual partners in the past 12 months previously mentioned in the interview: a current or former spouse, a current or former cohabiting partner, or a first sexual partner. Response category "20" for "None of the above" did not include all respondents who had a sexual partner in the past 12 months, but only those respondents who had ever married or cohabited with a partner but the sexual partner in the past 12 months was not previously mentioned. There are N=529 female respondents in 2013-2015 who were missing on P1YRELPx because they had more than 1 sexual partner in their lifetime, they had never married or cohabited, but had at least 1 sexual partner in the past 12 months who was not a previously mentioned partner. If users want to create a variable where P1YRELPx is applicable for all female respondents who had sexual intercourse in the past 12 months, they can reassign these cases from missing to code 20.

Section D: Computed variable ONLYTBVS

The purpose of the Blaise- computed variable ONLYTBVS is to indicate that the respondent's only sterilization operation is an unreversed tubal sterilization, or if she is currently married or cohabiting, only her husband's or partners unreversed vasectomy. Only those women who are coded "yes" on ONLYTBVS are routed into questions asking about their desire for sterilization reversal, either for their tubal or their husband's or partner's vasectomy. Due to a programming error in the instrument in 2013-2015, which has since been corrected in ongoing field interviewing, there were 226 unmarried/non-cohabiting female respondents who reported only an unreversed tubal sterilization, but were mistakenly set to "no" on ONLYTBVS instead of "yes." As a result of this miscoding, these 226 women did not get asked the question about their desire for reversal of their tubal sterilizations. On the public use file for female respondents, the values on ONLYTBVS were corrected to "yes" for all 226 women, and they were assigned the "not ascertained" code of 7 on DC-5 RWANTRVT, which they were not asked due to the programming/routing error.

Section E: Other-specify responses to EA-14 OTHRMETH (EA-15 SP_OTHRMETH)

EA_15 SP_OTHRMETH captures open-ended responses from respondents who chose "other method" ever used on EA-14 OTHRMETH. If the response was not among the methods already asked about in the EA series, it can be found in a new variable or variables (since EA-14 is 'Enter all that apply') created for this purpose: "NEWMETH." If a response indicated a method that was already asked about it was coded as "5" on NEWMETH: "Response was coded as a method in EA-1 through EA-14." Responses that appeared too infrequently to code separately are coded: "4" on NEWMETH: "Other method, not shown separately."

Section E: Open-ended questions for reasons for discontinuing specific contraceptives

The following variables in the EA series are a result of coding responses to open-ended questions (3 questions for each of 4 methods).

> STOPPILL1-STOPPILL6
> STOPCOND1-STOPCOND2
> STOPDEPO1-STOPDEPO5
> STOPIUD1-STOPIUD5

The open-ended follow-up questions were triggered based on certain responses to the following variables: EA-18 REASPILL, EA-19 REASCOND, EA-20 REASDEPO, EA-21a REASIUD. In cases where the verbatim response matched an already-existing category on one of these closed-ended questions, they were coded as that category on that variable. Verbatim responses that did not appear in sufficient numbers were collapsed into "Other side effects" and "Other – (Too few cases, not classifiable elsewhere)."

Section E: EA-21 TYPEIUD

The question was asked in 2011-2013 as well but only one response was allowed. Users who wish to combine 2011-2013 with 2013-2015 for this variable should note that it was changed to be a 'code all that apply", so responses are captured in both TYPEIUD_1 and TYPEIUD_2 in the 2013-2015 data.

Section E: ED Series, monthly contraceptive method history variables, METHX1-METHX192

Similar to past cycles, the ED series captures contraceptive method(s) the respondent used each month, for the time period from January, three years prior to the interview date, through interview date. For example, for interviews done in September 2013, the series collects method use information from January 2010 through September 2013. Variables for up to 4 methods for each month are saved on the data file. Therefore, variables containing these contraceptive methods range from "METHX1 through METHX192" (4 variables/methods per month, for up to 4 years, or 48 months.). With continuous interviewing, the number of calendar years spanned by the method calendar continues to increase. The output from this series was designed to remain the same number of month/year "cells," despite this accumulation of actual calendar years. Each respondent's particular window of 3+ years depends on the date they were interviewed. The following table displays the correspondence between calendar month and year and the ED contraceptive variable names and how this depends on date of interview, or cmintvw.

For convenience, two sets of frequently-used computed variables based on the method calendar are included in the female respondent data file.

currmeth1 - currmeth4: the method(s) the respondent used in the month of interview.
lstmonmeth1 - lstmonmeth4: the method(s) the respondent used in the month before interview.

If interview year is **2013** (cmintvw=**1365 (September)** through **1368**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2010	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2011	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2012	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2013	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

If interview year is **2014** (cmintvw=**1369** through **1380**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2011	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2012	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2013	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2014	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

If interview year is **2015** (cmintvw=**1381** through **1389 (September)**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2012	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2013	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2014	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2015	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

Section F: Mapping of family planning variables

In female Section F, women are asked about a set of 14 different possible services they could have received in the past 12 months. For each service received, the following information is gathered specific to that service: type of provider, whether that provider is the regular source of care, source of payment, etc. This table shows how the different variables are mapped by the type of service received.

Mapping of family planning service variables in 2013-2015 NSFG data release						
Service received in last 12 months	**Question number**	**Service variable name**	**Recodes**	**Provider**	**Regular source of care**	**Source of payment**
Method Birth Control/Prescription	FA-1b	BTHCON12	FPTITBC	BC12PLCX2	REGCAR12_F_02	BC12PAYX7-12
Checkup for Birth Control	FA-1c	MEDTST12	FPTITCHK	BC12PLCX 3	REGCAR12_F_03	BC12PAYX13-18
Counseling About Birth Control	FA-1d	BCCNS12	FPTITCBC	BC12PLCX 4	REGCAR12_F_04	BC12PAYX19-24
Sterilizing operation	FA-1e	STEROP12	FPTITSTE	BC12PLCX 5	REGCAR2_F_05	BC12PAYX25-30
Counseling re Getting Sterilized	FA-1f	STCNS12	FPTITCST	BC12PLCX 6	REGCAR12_F_06	BC12PAYX31-36
Emergency Contraception/Prescription	FA-1g	EMCON12	FPTITEC	BC12PLCX 7	REGCAR12_F_07	BC12PAYX37-42
Counseling re Emergency Contraception	FA-1h	ECCNS12	FPTITCEC	BC12PLCX 8	REGCAR12_F_08	BC12PAYX43-48
Pregnancy Test	FA-3a	PRGTST12	FPTITPRE	BC12PLCX 9	REGCAR12_F_09	BC12PAYX49-54
Abortion	FA-3b	ABORT12	FPTITABO	BC12PLCX 10	REGCAR12_F_10	BC12PAYX55-60
Pap Smear	FA-3c	PAP12	FPTITPAP	BC12PLCX 11	REGCAR22_F_11	BC12PAYX61-66
Pelvic Exam	FA-3d	PELVIC12	FPTITPEL	BC12PLCX 12	REGCAR2_F_12	BC12PAYX67-72
Prenatal Care	FA-3e	PRENAT12	FPTITPRN	BC12PLCX 13	REGCAR2_F_13	BC12PAYX73-78
Post-Pregnancy Care	FA-3f	PARTUM12	FPTITPPR	BC12PLCX 14	REGCAR2_F_14	BC12PAYX79-84
Counsel/test/treatment for STD	FA-3g	STDSVC12	FPTITSTD	BC12PLCX 15	REGCAR2_F_15	BC12PAYX85-90
All services in one visit:	FA-4	NUMBCVIS=1 all in one visit	-	BC12PLCX	REGCAR12_F_01	BC12PAYX- BC12PAYX 6

Section J (ACASI): JD-4 CONDVAG

There are 53 female teen respondents (aged 15-17) in 2013-2015 who had ever married, cohabited or had ever been pregnant (where VAGSEX was appropriately skipped and therefore sysmis) that were erroneously skipped past JD-4 CONDVAG. These cases were set to "not ascertained" for 2013-2015. This routing error was fixed for interviews beginning in January 2016.

Section J (ACASI): J-8b ANYORAL

The computed variable ANYORAL (computed in Flow Check J-8b) was set to "no" (coded "5") for 55 cases where <u>both</u> GIVORALM or GETORALM was equal to "refused" (coded "8") or "don't know" (coded "9"). Users should set these cases to missing so that only cases where both GIVORALM =5 AND GETORALM =5 are coded ANYORAL = 5. A similar issue occurred with the ANYORAL variable on the male file.

Female Recode LSEXRAGE: **Respondents with values of 45 years old**

On the female recode LSEXRAGE (respondent's age at last sexual intercourse), 6 respondents in 2013-2015 have a value of 45 years. One of these respondents were 44 years old at the time of the screener (computed variable agescrn=44), and had their 45th birthday before the interview was conducted (computed variable age_r and recode AGER=45), so LSEXRAGE is correctly equal to 45.

For the other 5 respondents, their age at last intercourse LSEXRAGE was actually 44, but was computed to be 45 based on the fact that last intercourse occurred in the month of interview, which happened to be their birth month. These respondents reported in question AA-1 AGE_A that they were 44 years old, indicating that the interview occurred <u>before</u> their actual birth<u>day</u> in that month. However, the formula for computing LSEXRAGE does not permit this level of precision because it is based only on century-months, not month/day/year. The formula yielded 45 years because it simply subtracts the century-month of the respondent's birth from the century-month of her last sexual intercourse and divides by 12 to obtain the age in years.

Female LSEXPAGE:

Since 2006-2010, this female file variable has been a computed variable instead of a recode. See 2011-2013 User's Guide Appendix 3a for the specifications of this variable that may assist you in constructing it for yourself with 2013-2015 data.

FEMALE PREGNANCY (INTERVAL) FILE NOTES

Section B (Pregnancy History) verification of key pregnancy data:

In the 2013-2015 NSFG, as in past NSFG surveys, all female respondents were asked questions about each of the pregnancies they have had, whether current or completed. They were asked to report these pregnancies in chronological order, starting with the first pregnancy, however in every survey period, some women did not report them in this order. Before moving on to questions that are based on pregnancies and pregnancy intervals in correct chronological order, it is critical that the pregnancy information is sorted manually (as in older Cycles before CAPI programming was introduced) or sorted by the programmed instrument (as in Cycle 5 in 1995).

In 2013-2015, a similar approach was taken as in 2006-2010 and 2011-2013, for sorting and correcting the pregnancy-specific information that women reported, and it was used to improve both the accuracy and completeness of the pregnancy data.

- After *each completed pregnancy loop*, women were asked to confirm key data about the pregnancy in the BH-6 CNFMPREG question – specifically, the pregnancy outcome, end date, and gestational length. The question wording for CNFMPREG was tailored to the type of outcome. For example, if the pregnancy ended in a single live birth, women were asked:

    "This pregnancy ended in the birth of 1 baby named *[first name or initials of baby]*. This pregnancy lasted *[X]* months and *[Y]* weeks and ended in *[baby's month/year of birth]*. Is this correct?

  If the woman said "yes," then the interviewer moved on the next pregnancy. If she said "no," she was given a chance to correct any information on outcome, end date, and gestational length.

- After all pregnancies were looped through the pregnancy-specific questions, women were asked to verify the chronological order of all of their pregnancies (BI-2 CHKORDER). Any current pregnancy was assumed to be the most recent pregnancy. All of the information for a woman's pregnancies are presented in a summary screen (see CAPI Reference Questionnaire for details), and she is presented with a summary statement about each pregnancy, similar to BH-6 CNFMPREG above. At this point, the respondent is given the following choices:

    - `Yes, pregnancies in order/everything is correct`
    - `No, pregnancies out of order`

  She may also volunteer that something else is incorrect; for example, she may decide to add or delete a pregnancy. If she reports that her pregnancies are out of order, she is asked BI-8 FIXORDER to place them all in the correct order. If any other key information about her pregnancies is incorrect, she is guided through specific questions to correct the information on outcome, end date, and gestational length.

  In any case, whether the respondent corrects any information or simply confirms that all is correct, **new variables** are created to hold the "corrected and chronologically sorted" pregnancy history data, or "**CCSD**" as the variable labels will indicate. These CCSD variables are passed forward to all later sections of the interview and are used in subsequent routing and consistency checks. The pregnancy

order variable (PREGORDR) and recode for birth order (BIRTHORD) on the public use file are both based on CCSD data.

The original variables as initially reported by the respondent _and_ the CCSD computed variables, which generally end in "_s," are included on the public use file. **For the vast majority of pregnancies reported by NSFG respondents, the CCSD variables are identical to the originally reported variables because no correction was needed.**

The table below provides a crosswalk of the original variables, the CCSD variables, and the corresponding recodes on the pregnancy or respondent file. All of the CCSD variables listed are defined in Flow Check B-42d (see CAPI Reference Questionnaire, Female Section B for details). The Flow Checks (FC) where the other computed variables are defined is listed in the table. As indicated throughout this User's Guide, it is best to use the official recodes when available, as they have had missing values imputed in a well-documented and consistent manner and will allow you to match your figures to published statistics. See **Appendix 3a (female respondent)** or **3b (pregnancy) recode specifications** for further details.

File	Short Description	Originally Reported Variable	Corrected and Chronologically Sorted Data (CCSD)	Recode
Resp	Number of pregnancies	BB-1 NUMPREGS	npregs_s	PREGNUM
Resp	Number of liveborn babies	numbabes (FC B-1)	nbabes_s	PARITY
Preg	Outcome of pregnancy	prgoutcome (FC B-42a)	outcom_s	OUTCOME
Preg	End date of pregnancy	cmprgend (FC B-42a)	cmendp_s	DATEND
Preg	Start date of pregnancy	cmprgbeg (FC B-42a)	cmpbeg_s	DATECON
Preg	Number of babies born alive from this pregnancy	bornaliv (FC B-10)	nbrnlv_s	n/a
Preg	Gestational length in weeks	wksgest (FC B-13)	n/a	PRGLNGTH

MALE FILE NOTES

Biological children in chronological order (Sections CG, DH, ED, FA, FC):

In the male questionnaire, respondents were asked about their biological children in the context of questions about their relationship with the mothers of these children. Based on consultation (before men were first included in the 2002 NSFG) with experts in surveys of men, this approach was considered to yield the most accurate reporting of men's sexual, fertility, and contraception experiences. However, with this approach, questionnaire information on a biological child is located on the data files in the section where a man reported the child:
– In Section C if the child's mother is his current wife or cohabiting partner
– In Section D if the child's mother was his last sexual partner ever, or one of his 3 most recent partners in the last 12 months
– In Section E if the child's mother was a former wife or his 1st cohabiting partner (and not a recent sexual partner)
– In Section F if the child's mother was any other sexual partner not discussed in Sections C-E

In each of the "woman-specific" loops in Sections C-E and in Section F, space was allowed for up to 10 biological children to be reported – for example, up to 10 children could be reported with the current wife or cohabiting partner in Section C, up to 10 with *each* former wife reported in Section E. In the 2013-2015 NSFG, there were never more than 9 children reported across all of these woman-specific loops in Sections C-E or in Section F. See exact numbers below for the reporting of biological children fathered across Section C-F among male respondents in the 2013-2015 NSFG:

- **Section C** – no respondent reported more than 7 children with his current wife or cohabiting partner

- **Section D** –
    - No more than 5 children reported with "P1" most recent partner
    - No more than 4 children reported with "P2" 2nd-most-recent partner in past year
    - No more than 3 children reported with "P3" 3rd-most-recent partner in past year

- **Section E** –no respondent reported more than 4 former wives were reported in the 2013-2015 NSFG and:
    - No more than 4 children reported with "W1" 1st former wife
    - No more than 2 children reported with "W2" 2nd former wife
    - No more than 2 children reported with "W3" 3rd former wife
    - No children reported with "W4" 4th former wife
    - No more than 5 children reported with "1st Cohab Partner"

- **Section F** – no respondent reported more than 9 biological children in this section

To assist users who wish to analyze information on men's biological children based on chronological order of birth, the male data file includes selected variables derived from the section-specific biological child variables in Sections C-F. These chronologically ordered variables are located on the male file at the end of Section F questionnaire items. Each of the characteristics listed below, with the exception of BIOLIVNGnn, are arranged as arrays of 10 variables each in the 2013-2015 NSFG, as no male respondent reported more than 10 biological children.

The chronologically-based BIO child variables included on male file are:

- BIODOBn: Century month of nth child's birth
- BIOSEXn: Sex of nth child
- BIOAGEn: Age of nth child
- BIOAGEGPn: Age group of nth child
- BIOHHn: Whether nth child lives in same household with R
- BIOMOMn: R's relationship to biological mother of this nth child *(this array is defined based on what section of the questionnaire the child was reported in)*
- BIOMARn: Whether R was married to nth child's mother at time of child's birth
- BIOCOHBn: Whether R was living with nth child's mother at time of child's birth (includes cohabiting or married)
- BIOLRNPGn: When R learned of the pregnancy (before or after child was born)
- BIOLIVNGnn: Where child is living now (4 mentions for nth child, eg., biolivng11 indicates the 1st mention for R's 1st biological child; biolvng24 indicates the 4th mention for R's 2nd biological child)
- BIOCHSIGn: If nth child born outside marriage: Whether R signed application for birth certificate or other paternity document
- BIOCHCRTn: If nth child born outside marriage: Whether R went to court to establish paternity
- BIOCHGENn: If nth child born outside marriage: Whether R established paternity by blood or genetic test
- BIOLVEVRn: Whether R ever lived with nth child (if not living with child now)
- BIOHWFARn: How far away nth child lives (in miles) from respondent
- BIOWANTn: For nth child: Wantedness of the pregnancy by R right before it began
- BIOHSOONn: For nth child: Timing of the pregnancy right before it began
- BIOHPYPGn: For nth child: R's happiness about the pregnancy when he learned about it

The table on the next page illustrates how these chronologically arranged variables are derived from questionnaire items in Sections C-F. Please consult the codebook to see further details on universe statements and response categories for these variables or the male questionnaire CRQs for the precise wording and routing for each item in Sections C-F.

In summary:
These chronologically ordered variables are essentially identical in content to the source variables in Sections C-F. They are provided for the user's convenience and hold no different values than the originally collected data in Sections C-F; they only arrange the information in a different (chronological) manner. Users whose primary goal is to examine data on men's biological children in order of their birth may prefer to use these chronologically arranged variables. If, however, their primary goal is to examine men's fertility in the context of their relationships with their children's mothers and they wish to incorporate other variables describing those relationships, it may be easier or more appropriate to use the source variables in Sections C-F.

Chronologically arranged BIO variable	Brief Description	Section C source	Section D source	Section E source	Section F source
BIODOB[x]	Century Month date of birth	CG-6 CWPCHDOB_M/Y yields cmchdob[x]	DH-6 PXCXBORN_M/Y yields cmchdob[x]	ED-6 FWPCHDOB_M/Y yields cmchdob[x]	FA-9m/y OBCDOB_M/Y yields cmchdob[x]
BIOSEX[x]	Sex of child	CG-5 CWPCHSEX	DH-5 PXCXSEX	ED-5 FWPCHSEX	FA-8 OBCSEXX
BIOAGE[x]	Age of child in years	based on cmchdob[x]	based on cmchdob[x]	based on cmchdob[x]	based on cmchdob[x]
BIOAGEGP[x]	Age group of child	CG-12 CWPCHAGE	DH-12 PXCXAGE	ED-12 FWPCHAGE	FA-15 OBCAGE
BIOHH[x]	Is child in R's household	based on CG-11 CWPCHLIV	based on DH-11 PXCXLIV	based on ED-11 FWPCHLIV	based on FA-14 OBCLIV
BIOMOM[x]	Relationship of child's mother to R (coded based on which section child was reported in)	set to 1 if current wife; set to 2 if current cohab	set to 3 if recent or last partner & also a former wife/cohab partner; set to 4 if any other recent or last partner	set to 5 if former wife; set to 6 if 1st cohab (but not a former wife)	set to 7 if child was reported in this section or not otherwise coded
BIOMAR[x]	Was R married to mother at time of child's birth	CG-8 CWPCHMAR	DH-8 PXCXMARB	ED-8 FWCHMARB	Inapp (all nonmar)
BIOCOHB[x]	Was R living with mother at time of child's birth (married or cohabiting)	CG-8 CWPCHMAR & CG-9 CWPCHRES	DH-8 PXCXMARB & DH-9 PXCXRES	ED-8 FWCHMARB & ED-9 FWPCHRES	FA-12 OBCMLIV
BIOLRNPG[x]	When R learned of the pregnancy	CG-10 CWPCHLRN	DH-10 PXCXKNOW	ED-10 FWPCHLRN	FA-13 OBCKNOWX
BIOLIVNG[x]	Where child living now (up to 4 mentions for each child)	CG-11 CWPCHLIV	DH-11 PXCXLIV	ED-11 FWPCHLIV	FA-14 OBCLIV
BIOCHSIG[x]	Whether established paternity by signing birth certificate application	CG-13a CWPCHSIG	DH-13a PXCXSIG	ED-13a FWPCHSIG	FA-16A OBCCHSIG
BIOCHCRT[x]	Whether established paternity by going to court	CG-13b CWPCHCRT	DH-13b PXCXCRT	ED-13b FWPCHCRT	FA-16b OBCCHCRT
BIOCHGEN[x]	Whether established paternity by blood or genetic test	CG-14 CWPCHGEN	DH-14 PXCXGEN	ED-14 FWPCHGEN	FA-17 OBCCHGEN
BIOLVEVR[x]	Has R ever lived with child (if not living with child now)	CG-15 CWPCHEVR	DH-15 PXCXEVER	ED-15 FWPCHEVR	FA-18 OBCEVER
BIOHWFAR[x]	How far away does child live (miles)	CG-16 CWPCHFAR	DH-16 PXCXFAR	ED-16 FWPCHFAR	FA-19 OBCFAR
BIOWANT[x]	Wantedness of this pregnancy	CG-17 CWPCHWNT	DH-17 PXWANT	ED-17 FWPRWANT	FA-20 OBCRWANX
BIOHSOON[x]	Timing of this pregnancy	CG-18 CWPCHSON	DH-18 PXSOON	ED-18 FWPSOON	FA-21 OBCSOONX
BIOHPYPG[x]	R's happiness about this pregnancy	CG-19 CWPCHHPY	DH-19 PXHPYPG	ED-19 FWPHPYPG	FA-22 OBCHPYX

Section K (ACASI): ANYORAL

The computed variable ANYORAL (computed in Flow Check K-6c) was set to "no" (coded "5") for 33 cases where both GIVORALF or GETORALF was equal to "refused" (coded "8") or "don't know" (coded "9"). Users should set these cases to missing so that only cases where both GIVORALF =5 AND GETORALF =5 are coded ANYORAL = 5. A similar issue occurred with the ANYORAL variable on the female file.

Recode LSEXRAGE (based on data from sections B-D): Respondents with values of 45 years old

On the male recode LSEXRAGE (respondent's age at last sexual intercourse), 4 respondents in 2013-2015 have a value of 45 years. One of these respondents were 44 years old at the time of the screener (computed variable agescrn=44), and had his 45th birthday before the interview was conducted (computed variable age_r and recode AGER=45), so LSEXRAGE is correctly equal to 45.

For the other 3 respondents, their age at last intercourse LSEXRAGE was actually 44, but was computed to be 45 based on the fact that last intercourse occurred in the month of interview, which happened to be their birth month. These respondents reported in question AA-1 AGE_A that they were 44 years old, indicating that the interview occurred before their actual birthday in that month. However, the formula for computing LSEXRAGE does not permit this level of precision because it is based only on century-months, not month/day/year. The formula yielded 45 years because it simply subtracts the century-month of the respondent's birth from the century-month of her last sexual intercourse and divides by 12 to obtain the age in years.