Public Use Data File Documentation

2006-2010

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 2011

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. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information.

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

What You Need to Know about This User's Guide for the 2006-2010 NSFG

This User's Guide is for researchers planning to use the 2006-2010 National Survey of Family Growth (NSFG) Public Use Data Files. This version has been updated with new information since the release of the 2006-2008 Public Use Data Files and **replaces** the previous User's Guide. It is intended to guide you through the process of identifying the variables you need for your research. The User's Guide consists of:

- **Part 1** provides an overview with information that applies to all users, regardless of their research topics
- **Part 2** has information about specific topics relevant to your analysis
- **7 Appendices** that provide reference material:
  - Appendix 1 – Indexes to all the variables on the public use data files
  - Appendix 2 – SAS and Stata syntax for common file manipulations
  - Appendix 3 – Specifications for all the Recode (constructed) variables
  - Appendix 4 – Crosswalk grids for all Recode variables and their equivalents by gender and across NSFG survey years
  - Appendix 5 – Summary of questionnaire changes in the 2006-2010 NSFG
  - Appendix 6 – Frequently Asked Questions (FAQ) for the NSFG
  - Appendix 7 – Restricted-Use variables from the 2006-2010 NSFG that are available through the NCHS Research Data Center (not including the contextual data)

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:

1. **Use sample weights (e.g., WGTQ1Q16) and design variables (SECU & SEST)** to make valid estimates. Failure to use the weights and design variables correctly will lead to inaccurate findings.
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 24.
3. In addition to this User's Guide with its 7 appendices, the **NSFG website** also provides:
   a. the **questionnaires** in 2 formats
   b. the **codebooks** (online webdoc and freestanding PDF files)
   c. downloadable **ASCII data files**
   d. **program statements** to read in the ASCII data using SAS, Stata, and SPSS
   e. additional examples of variance estimation
4. If you cannot find an answer to your question in the NSFG User's Guide,

codebooks, questionnaire documents, or webpage, contact NSFG staff at nsfg@cdc.gov or 301-458-4222.

5. Refer to the **two technical reports** (Series 1, Number 48 and Series 2, Number 150) for detailed information on the planning, sample design, operations, or statistical aspects of the 2006-2010 NSFG. Another Series 2 report is expected to be released in 2012. See the draft summary of that report on pages 13-17.

For published reports or statistics based on NSFG data, please see the following:

- The NSFG website (www.cdc.gov/nchs/nsfg.htm) includes lists of **reports and articles using all Cycles of the NSFG**. NSFG reports published by the National Center for Health Statistics (NCHS) are accessible for download in PDF format from the NSFG website. Single copies of printed reports are available for free, as long as supplies last, by emailing nsfg@cdc.gov,
- **Another source** of published NSFG data is "**Key Statistics**," accessible on the NSFG web page (http://www.cdc.gov/nchs/nsfg/abc_list.htm). **Key Statistics** briefly present one or more of the most important statistics from nearly all topics covered in the NSFG. These statistics were originally based on Cycle 6 of the National Survey of Family Growth (NSFG), conducted in 2002. As reports are published based on data from the 2006-2010 NSFG, statistics will continue to be updated.

USER'S GUIDE PART 1: GENERAL INFORMATION FOR USERS OF THE 2006-2010 NSFG PUBLIC USE DATA

As noted above, the User's Guide for the 2006-2010 NSFG is divided into 2 parts.

Part 1 provides general information for users of the public use data, including:

- Overview of the 2006-2010 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 7 appendices:

1. File indexes (i.e., listings of each variable with a brief description)
2. SAS & Stata syntax guidelines for common file manipulations
3. Recode specifications for specially constructed and imputed variables
4. Recode cross-walks by gender and across years of data collection
5. Summary of NSFG questionnaire changes from Cycle 6 (2002) and across years of the 2006-2010 NSFG
6. Frequently Asked Questions about the NSFG
7. List of restricted-use variables from the 2006-2010 NSFG interviews that are available through the NCHS Research Data Center

BACKGROUND AND OVERVIEW OF THE 2006-2010 NSFG

The National Survey of Family Growth (NSFG) is designed and administered by the National Center for Health Statistics (NCHS), an agency of the US Department of Health and Human Services, in collaboration with several other federal agencies (see Acknowledgments). The NSFG has been conducted 7 times since 1973. The purpose of the survey is to produce 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 adoption;
- adoption and caring for nonbiological children
- father involvement behaviors and
- men's and women's attitudes about sex, childbearing, and marriage.

The survey results are used by the U.S. Department of Health and Human Services and other research and policy organizations to plan health services and health education programs, and to do statistical studies on the topics listed above.

For the 2006-2010 NSFG, statistical design, interviewing, and data processing have been conducted by the **University of Michigan's Institute for Social Research (ISR)**, under a contract with the National Center for Health Statistics (NCHS), in collaboration with the NCHS NSFG team led by William Mosher (Project Officer).

The following table presents basic information on each of the previous cycles of the NSFG, and on the first release of 2006-2010 interviewing, the focus of this User's Guide.

Cycle	Year	Scope	Number of Interviews	Over-Samples	Average Length	Incentive
1	1973	Ever-Married Women 15-44	9,797	Black Women	60 Minutes	No
2	1976	Ever-Married Women 15-44	8,611	Black Women	60 Minutes	No
3	1982	All Women 15-44	7,969	Black Women Teens	60 Minutes	No
4	1988	Women 15-44	8,450	Black Women	70 Minutes	No
5	1995	Women 15-44	10,847	Black Women Hispanic Women	100 Minutes	$20
6	2002	Women 15-44 Men 15-44 (First time)	12,571 W = 7,643 M = 4,928	Blacks Hispanics Ages 15-24	W= 85 min M= 60 min	$40
n/a	2006-2010	Women 15-44 Men 15-44	22,682 W = 12,279 M = 10,403	Blacks Hispanics Ages 15-24	W=80 min M=60 min	$40

Interviewing for the release of the 2006-2010 NSFG was conducted from June 2006 through June 2010. In-person interviews were conducted with 12,279 women 15-44 years of age and 10,403 men 15-44 years of age for a total sample size of 22,682. The interviews were conducted by trained female interviewers using laptop, or notebook, computers---a procedure called computer-assisted personal interviewing (CAPI). The interviews for women averaged 80 minutes; the interviews for men averaged 60 minutes. The response rate was 77% overall—78% for females, 75% for males, and 77% for male and female teenagers.

Additional details on how the survey was designed and conducted may be found on the NSFG web site (www.cdc.gov/nchs/nsfg.htm), as well as in the following technical reports:

RM Groves et al. Planning and Development of the Continuous National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 48. September 2009. Hyattsville, MD: National Center for Health Statistics. Available at http://www.cdc.gov/nchs/data/series/sr_01/sr01_048.pdf.

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

JM Lepkowski et al. Results of Fieldwork, Weighting,Imputation, and Variance Estimation in the 2006-2010 National Survey of Family Growth. *Vital and Health Statistics*, Series 2, forthcoming 2012.

HOW TO OBTAIN PUBLIC USE DATA AND DOCUMENTATION and OTHER DATA FILES

The public use data and documentation for the 2006-2010 NSFG are available on the NSFG website (www.cdc.gov/nchs/nsfg.htm). Documentation includes this User's Guide, along with codebook documentation for each variable on the files, and questionnaires for each survey year.

The public use data for the 2006-2010 NSFG consist of 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)

In addition to the public use files in from the 2006-2010 NSFG, the following files for 2006-2010 will become available within the next year.

- ACASI file
- Contextual Data file
- Interviewer Observation file

ACASI file: This file contains most of the survey data collected using Audio Computer-Assisted Self Interviewing (ACASI), in which the respondent enters the answers directly into the computer without the help of an interviewer. The object of ACASI was to give respondents a more private opportunity to report sensitive information. See "Outline of Contents of the Data

Files" and the questionnaires available on the NSFG website for more information on the specific items included in the ACASI section. Essentially the items describe current and past behavior related to the risk of acquiring sexually transmitted diseases (STD) including the Human Immunodeficiency Virus, or HIV, the virus that causes AIDS. The ACASI file is provided to researchers free of charge, upon request. Given the sensitivity of the data items, however, a user agreement must be signed before gaining access to this file.

Contextual Data file: The contextual data files include restricted-use information on the context, or community environment, in which respondents live (for example, the unemployment rate in the area, or the percent of households with incomes below the poverty level). As in Cycles 5 and 6, a contextual data file will be available for the 2006-2010 NSFG, through the NCHS Research Data Center, given the risk of disclosure posed by these geographic variables. Contextual variables will be based, to the extent possible, on where the respondents lived at 2 points in time – April 1, 2000 (the time of the last U.S. Census) and the date of interview for the 2006-2010 NSFG. For further information on the NSFG contextual data, you may wish to consult the NCHS Research Data Center website: http://www.cdc.gov/rdc/.

ORGANIZATION OF THE 2006-2010 NSFG PUBLIC USE DATA FILES

The public use data for the 2006-2010 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 = 2006_2010_FemResp.dat (one record per woman)	12,279	6,251	3,741
Female pregnancy (interval) file File = 2006_2010_FemPreg.dat (one record per pregnancy)	20,497	562	287
Male respondent file File = 2006_2010_Male.dat (one record per man)	10,403	4,543	2801

The **Female Respondent file** contains one record for each of the 12,279 women in the survey and includes most of the information from their interviews. The **Female Pregnancy (Interval) file** contains one record for each of 20,497 pregnancies (both completed pregnancies

and current pregnancies), and contains information about the characteristics of each pregnancy and method use and wantedness before each 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 third data file, the **Male Respondent** file, often just referred to as the male file, contains one record for each of the 10,403 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. 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 column locations of the major sections and key variables contained in all 3 public use files from the 2006-2010 NSFG.

- For more detail on data file layout than what is shown below, see **Appendix 1** (File Indexes) in which column locations and short descriptions are provided for every variable included on the public use files.
- More detailed outlines of the male and female questionnaires are provided in Figures 6 and 7 in the Series 1 Report:

  RM Groves et al. Planning and development of the Continuous National Survey of Family Growth, 2006-2008. *Vital and Health Statistics*, Series 1. No. 48. Hyattsville, MD: National Center for Health Statistics. September 2009. (www.cdc.gov/nchs/data/series/sr_01/sr01_048.pdf)
- Though the male and female questionnaires were organized somewhat differently, they included much of the same information. **Appendix 4** provides a "crosswalk" of recodes for males and females.

FEMALE RESPONDENT FILE *– information for each female*

Beginning Column #	Items or Sections
1	Respondent ID (CASEID) and selected screener variables
11	Questionnaire Data (including computed variables) for Sections A-J
11	A: Background and demographic information
98	B: Pregnancy and adoption-related information
727	C: Marital and relationship history; first sexual intercourse; recent partners
1706	D: Sterilizing operations and impaired fecundity
1904	E: Contraceptive history and related information
4391	F: Family planning and medical services
4539	G: Birth desires and intentions
4558	H: Infertility services and reproductive health
4696	I: More background, more demographic information, & attitude questions
4815	*J: Audio CASI: general health measures; pregnancy reporting; drug use; STD/HIV-risk behaviors; nonvoluntary intercourse; income

4853	Recodes (constructed, imputed variables) & imputation flags (including key recodes describing pregnancies)
6132	Weights & related variables *(including sample design variables)*
6226	Date of interview, questionnaire version, and related variables

* Only income and selected health measures from Audio CASI are included on this first public use file release. The remainder of variables from ACASI will be included on the ACASI files to be released within the year.

FEMALE PREGNANCY (INTERVAL) FILE *– information for each pregnancy (including current)*

Beginning Column #	Items or Sections
1	Respondent ID (CASEID)
6	Pregnancy Order (PREGORDR)
8	Questionnaire Data (including computed variables) for Sections B & E
8	B: pregnancy outcomes and dates, prenatal care, sources of payment for delivery, maternity leave, breast-feeding
261	E: contraceptive use in the pregnancy interval and wantedness of the pregnancy
331	Recodes (constructed, imputed variables) & imputation flag (including key recodes and other variables from respondent file)
449	Weights & related variables *(including sample design variables)*
543	Date of interview, questionnaire version, and related variables

MALE RESPONDENT FILE *– Information for each male*

Beginning Column #	Items or Sections
1	Respondent ID (CASEID) and selected screener variables
12	Questionnaire Data (including computed variables) for Sections A-K
12	A: Background and demographic information
107	B: Sex education, vasectomy & infertility, sexual intercourse, enumeration and relationship with up to 3 recent (or last) sexual partners
229	C: Current wife or cohabiting partner: date of marriage; children; contraception with her
867	D: Recent (or last) sexual partner(s) (up to three): key dates, children; contraception with her; 1st partner
1954	E: Former wives and first premarital cohabiting partner: key dates, children; contraception with each
3034	F: Other biological and adopted children; other pregnancies
3691	G: Fathering: Activities with R's children – those he lives with and those he doesn't live with
3745	H: Desires and intentions for future children
3762	I: Health conditions, access to health care, and receipt of health services

3880	J: More background, more demographic information, & attitude questions
3969	*K: Audio CASI: pregnancy reporting; drug use; STD/HIV-risk behaviors; nonvoluntary intercourse; income
4007	Recodes (constructed, imputed variables) & imputation flags
4428	Weights & related variables *(including sample design variables)*
4522	Date of interview, questionnaire version, and related variables

* Only income and selected health measures from Audio CASI are included on this first public use file release. The remainder of variables from ACASI will be included on the ACASI files to be released within the year.

SAMPLE WEIGHTS AND VARIANCE ESTIMATION

These 2006-2010 NSFG data are based on a nationally representative sample of the household population aged 15-44. Data users should understand how to account for the design when doing their analyses. This section provides a summary of the procedures used for sample weighting and variance estimation. The 2006-2010 NSFG design was a significant change in the design of the survey. Thus, two reports were released that described how the survey was planned and designed. Full background on these topics, along with sample design for the NSFG, can be found in these two reports, along with a third report forthcoming in 2012:

- RM Groves et al. Planning and Development of the Continuous National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 48. September 2009. Hyattsville, MD: National Center for Health Statistics. Available at http://www.cdc.gov/nchs/data/series/sr_01/sr01_048.pdf.
- JM Lepkowski et al. Continuous National Survey of Family Growth: Sample Design, Sampling Weights, Imputation, and Variance Estimation, 2006-2008 *Vital and Health Statistics*, Series 2. No. 150. June 2010. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf
- JM Lepkowski et al. Results of Fieldwork, Weighting,Imputation, and Variance Estimation in the 2006-2010 National Survey of Family Growth. *Vital and Health Statistics*, Series 2, forthcoming 2012.

Sampling Weights

In the NSFG, Hispanics, blacks, teens, and women were selected at higher rates than others in the 15-44 age group. "Sampling weights" adjust for these different sampling rates, as well as for response rates and coverage rates, so that accurate national estimates can be made from the sample. Because weights vary widely for 2006-2010 NSFG respondents, we strongly recommend using the weights in all analyses. Using the weights will also permit replication of the nationally representative estimates that appear in published NCHS reports.

There are five weights provided on the 2006-2010 public use files. The weight WGTQ1Q16 is the main weight to be used in most analyses involving all 16 quarters of data. The 4 additional weights are provided for analyses of questions that were added at different times in data collection or to compare cases that were collected in the first 2 years of data collection (quarters 1-8) with cases collected in the last 2 years of data collection (quarters 9-16):

(a): **WGTQ1Q16:**
This weight applies to variables available for all 4 years of data collection (quarters 1 to 16): June 2006 through June 2010. This is the weight that is most analogous to previous cycles of the NSFG. It should be used for most analyses.

(b): **FINALWGT30:**
This is the same weight that was provided with the 2006-2008 public use files. It is the weight used when analyzing the 1st 2 1/2 years (or 30 months) of data collection (quarters 1 to 10). It is applicable only for interviews conducted from June 2006 through December 2008.

(c): **WGTQ1Q8:**
This weight is applicable only for cases interviewed in years 1 and 2 of data collection (quarters 1 to 8, June 2006 through June 2008). It should be used when analyzing those cases relative to those in quarters 9-16.

(d): **WGTQ5Q16:**
This weight is applicable only for cases interviewed in years 2, 3, and 4 of data collection (quarters 5 to 16). This weight would typically be used for analyzing items that were added in **Year 2** (quarter 5), beginning in July 2007.

(e): **WGTQ9Q16:**
This weight is applicable only for cases interviewed in years 3 and 4 of data collection (quarters 9 to 16). This weight would typically be used for analyzing items that were added in **Year 3** (quarter 9), beginning in July 2008.

Sources of information on questions that were added or enhanced in Years 2 and 3 include:

- **Codebook documentation for the variable**: For any variable in which a change was made in Year 2 or Year 3, the variable **QuestYear** is referenced QuestYear takes the values of 1, 2, and 3 – these correspond to the years of data collection. Note that a value of 3 represents years 3 and 4 because no questionnaire changes were introduced in year 4. QuestYear identifies in which year questionnaire changes occurred. It is referenced in the question text if question wording changed from one year to the next. It appears in the Universe Statement if the variable was applicable for different groups of cases in different years. QuestYear may also be referenced in the value label of a variable to indicate changes in response categories across questionnaire years. In such cases, both sets of response categories, pre- and post-change, are generally presented.
- **Appendix 5** of the User's Guide provides a summary of all changes to the questionnaire including added questions. This section refers you to the codebook entries, questionnaires, or User's Guide as appropriate for further details.
- **Questionnaires** for Years 1, 2, and 3/4 are provided on the NSFG webpage as well. For some questionnaire changes that affect whole series of questions, the questionnaires may be the best way to understand the changes that were made and any impact on your analyses.

All five weights are included on the Female Respondent file, the Female Pregnancy file, and the Male file. They are always located near the end of the data record, and their exact column

locations can be found in **Appendix 1a-1c** (File Indexes) or in the codebook.

Each weight yields national totals for the population 15-44 (about 125 million men and women in the US, aged 15-44 in 2006-2010). WGTQ1Q16, the 4-year weight, yields estimates for these 125 million men and women based on 22,682 interviews**.**

To yield number in thousands, as often appears in NCHS reports, you divide each sample weight by 1,000. For example, if analyzing the entire 4 years of data, you would use this data manipulation step to create a new weight variable:

WGT1000= WGTQ1Q16/1000

Variance Estimation

Sampling variance is a measure of the quality of a statistic (such as a percentage, proportion or a mean) caused by having taken a sample instead of interviewing the full population. In the 2006-2010 NSFG, the sampling variance measures variation caused by interviewing the NSFG sample instead of all women and men 15-44 in the entire country.

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. Special software is required to accurately estimate sampling errors in a complex sample such as the NSFG; such software is increasingly common and easy to use.

For example, SAS procedures such as "FREQ" produce population variances assuming simple random sampling, but SAS has also developed procedures for complex survey estimates in its 'SURVEY' procedures such as "SURVEYFREQ." Similarly, Stata and SPSS have the Taylor series procedures for complex surveys in 'svy' commands or a complex sample survey module, respectively.

Examples of how to use such software to estimate sampling errors for Cycle 6 (2002) are posted on the NSFG web site at http://www.cdc.gov./nchs/nsfg/nsfg_cycle6.htm, and further guidance for 2006-2010 will be posted when available. In addition, selected examples are provided in **Appendix 2**.

For example, the sample design parameters needed to estimate variances that account for the complex sample design of the 2006-2010 NSFG using SAS/SUDAAN statistical software are:

DESIGN = WR (with replacement)

NEST statement:
SEST = stratum variable
SECU = PSU (cluster) variable

WEIGHT statement:
WGTQ1Q16 = final post-stratified, fully adjusted weight for 4 years.

Here is a specific example (based on SUDAAN version 9 code) for a tabulation of the recode HADSEX from the male data file, using the DEFF option to calculate design effects:

```
proc sort data=nsfgmale;
  by SEST SECU;
  run;
proc crosstab data=nsfgmale design=wr deff;
  nest SEST SECU;
  weight WGTQ1Q16;
  class hadsex;
  table hadsex;
  print nsum wsum rowper serow deffrow;
run;
```

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 use the *full* dataset to set up your analytic variables, and define your subgroup of interest within your procedure. This minimizes the chances of encountering empty cells across the sample clusters or strata that could cause errors in your statistical software packages. Using the appropriate statement for your software, you would then subset cases within your procedure, using variables already in the file or new variables you have created. In SUDAAN, for example, you would use the SUBPOPN statement to subset your cases of interest. See example below that subsets cases younger than age 25 using the recode AGER:

```
proc crosstab data=nsfgmale design=wr deff; /* full dataset used here */
  nest SEST SECU;
  weight WGTQ1Q16;
  subpopn ager < 25;  /* aubset cases who are < 25 years old */
  class hadsex;
  table hadsex;
  print nsum wsum rowper serow deffrow;
run
```

A Note on Sampling Errors in the 2006-2010 NSFG

Sampling errors in complex survey sample designs are affected by a number of factors, only one of which is sample size. The increases in sample size between the 2002 NSFG (n= 12,571) and the 2006-10 NSFG (n = 22,682) did not always result in a proportionate reduction in the variances of the statistics from the NSFG sample. The effects are best illustrated with examples that estimate comparable proportions using comparable denominators from the 2002 and 2006-10 surveys (See section in Part 2 of this User's Guide).

The effects of all these changes in the sample design will vary from one statistic and one sub-group to another. But in general, the gains in precision from the 2002 survey to the 2006-2010 survey were greatest in the over-sampled groups-- including Black, Hispanic, and teenage respondents. For white females and for all females, standard errors did not always shrink from 2002 to 2006-2010, in spite of the larger sample size. That is because of several design features in the survey:

- First, the sample sizes in 2006-2010 increased by about 60% for females, from 7,643 in 2002 to 12,279 in 2006-2010. But the sample sizes for males increased by more than 100%, from 4,928 in 2002 to 10,403 in 2006-2010. This was a deliberate decision to increase the proportion of males in the sample from 40% males (and 60% females) to 45% males (and 55% females). Therefore there was a more consistent, and larger, reduction in standard errors for men than for women.

- Second, the sample in 2006-2010 was more clustered than in 2002. Clustered samples are less expensive to work, because the interviewer spends less time traveling to visit respondents repeatedly to find people at home and to conduct interviews, and because fewer interviewers are needed. Since cases that are in a "cluster" (i.e. cases that are close together such as in the same neighborhood) tend to be more similar to each other than cases far away, this more clustered sample has higher sampling errors than a simple random sample of the same size.

- Third, because of the need to obtain large samples of black, Hispanic and teenage respondents, there was a great deal of variation in the sampling weights (these weights tell the analyst how many people in the population are represented by a person in the sample).

- Fourth, there was variation in the weights because non-respondents were sub-sampled in the non-response follow-up. About 8.5% of the final sample was obtained in the non-response follow-up, and these cases have much larger weights than other cases.

Users of the NSFG 2006-2010 data files will find that the standard errors of some statistics in the NSFG may not be as small as they may expect. These standard errors were the result of a successful effort to produce large samples of teens, blacks, and Hispanics at an affordable cost.

OVERVIEW OF DATA QUALITY IN THE NSFG

As measured by amounts of missing data and inconsistent data, data quality in the 2006-2010 NSFG is high, as it was in previous cycles. This high quality was obtained through:

-- questionnaire design work, including specification, testing, and the experience of Cycle 6 (2002);
-- consistency checks built into the interview that allow more 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.

Among the thousands of variables in the file, Part 2 of the User's Guide provides information on

(a) selected questionnaire items or recodes, or
(b) variables that have questionable or "not ascertained" responses. The issue of questionable or not ascertained data affects only a small percent of cases. Knowing these issues, however, allows users to exclude those cases, or deal with them in some other way.

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 assistance apart from what is discussed below, 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 2006-2010 NSFG (as in Cycle 6) 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 2006-2010 NSFG, the process of checking for consistency by NCHS and ISR staff 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, given the size and complexity of these data files, they are not 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 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 in Section I (females) and Section J (males). The small number of respondents who did not complete the ACASI section, partially or completely, will have "not ascertained" values from 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 codebook documentation only shows these codes if the variable has cases with 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."

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 recorded as month and year, except for the respondent's date of birth, which was recorded as month, day, and year.

For inclusion in the data file, month and year for most dates reported in the interview, including the respondent's date of birth, were converted to "century months" by subtracting 1900 from the year, then multiplying the remainder by 12, and adding the number of the month, where January = 1, February = 2, and so on.

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 that 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 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 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 277 (January 1923) through 1326 (June 2010) are shown in the array below with the years from 1923 through 2010 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 2006-2010 NSFG were conducted between June 2006 (century month 1278) and June 2010 (century month 1326).

DATE CODES

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1923	277	278	279	280	281	282	283	284	285	286	287	288
1924	289	290	291	292	293	294	295	296	297	298	299	300
1925	301	302	303	304	305	306	307	308	309	310	311	312
1926	313	314	315	316	317	318	319	320	321	322	323	324
1927	325	326	327	328	329	330	331	332	333	334	335	336
1928	337	338	339	340	341	342	343	344	345	346	347	348
1929	349	350	351	352	353	354	355	356	357	358	359	360
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

Other-Specify Coding

In the 2006-2010 NSFG, 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 CRQs with a "_SP" at the beginning or end of their name, and the question text and response type distinguish these clearly as open-ended questions. In all cases, responses that were clearly examples of a pre-existing category (part of the questionnaire instrument and shown to respondents on a showcard) were recoded (or "back-coded") into that pre-existing category. The remaining unique responses were examined by NCHS staff and if the numbers were sufficient, the new response categories were added to the list of existing (pre-specified) response categories for subsequent interviews. In a few instances, the responses were categorized and a new variable was created that contained only these additional responses.

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. The actual verbatim responses are not included in the public use data files, but they have all been reflected in some way in the associated, closed-ended variables 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).

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, 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, or 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 codebook pages** stating the name of the 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:

- recodes for males and females in 2006-2010 NSFG (arranged by topics)
- recodes for females across 1995, 2002, and 2006-2010 NSFG (arranged by section of questionnaire)
- recodes for males across 2002 and 2006-2010 NSFG (arranged by section of questionnaire)

The frequency of missing values for the recoded variables in 2006-2010 is quite low, just as it was in the 2002 NSFG (Cycle 6). 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.

Relatively few cases or recodes were imputed using logical imputation, which involves having a subject-matter expert at NCHS examine variables related to the variable in question, and assign a value that is consistent with those other variables and is an educated guess of the

true value when there is any ambiguity. Logical imputation is also assigned as the flag value when a recode's value can be defined based on the imputed values of source recodes. Most recode values were assigned using regression imputation software in which multiple regression is used to predict a value for the case using all other variables in the data set as predictors. Regression imputation follows the same logical constraints that are built into the original recode's 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.

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 record which kind of imputation was used. Each imputation flag has the following potential values:

0=Questionnaire data (no imputation)
1=Multiple regression imputation (used most often)
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. For further details on the imputation process for the recodes, consult the Series 2 report:

> JM Lepkowski et al. Continuous National Survey of Family Growth: Sample Design, Sampling Weights, Imputation, and Variance Estimation, 2006-2008 *Vital and Health Statistics*, Series 2. No. 150. June 2010. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_150.pdf

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. **We recommend that analysts use all, including the imputed, cases**. 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.

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. **Appendix 3** of the User's Guide contains the full specifications for all recodes, roughly in order of the questionnaire sections on which they are based. **Appendix 4** provides recode crosswalks to assist you in finding comparable recodes across NSFG surveys or for males and females.

Protections to Minimize Risk of Disclosure for Individual-Level Data

When NCHS collected data from respondents to 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. In order to do so, some modifications were made 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.

The public use files were reviewed by the NCHS Disclosure Review Board and the NCHS Confidentiality Officer. In response to that review, the NSFG staff and contractor made changes to minimize the risk of disclosure, including the suppression or collapsing of additional variables that could be used to identify very small groups.

All directly identifying information (names, addresses) has been eliminated from the public use files. In addition, the state and Census region of residence have been withheld. The only geographic variable included on the public use files is a 3-category METRO recode (metro-central city, metro-suburb, or non-metro). 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.

As a final step to prevent identification of individual respondents, the values of some variables have been altered for some respondents. That is, a few values in the data set are not the actual values reported by the respondents. 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.

Duncan GT, Stokes SL. 2004. Disclosure Risk vs. Data Utility: The R-U Confidentiality Map as Applied to Topcoding. Chance 17(3):16-20.

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

Feinberg, SE. 2001. Statistical Perspectives on Confidentiality and Data Access in Public Health. Statistics in Medicine 20(9-10):1347-56.

Feinberg SE, McIntyre J. 2004. Data Swapping: Variations on a Theme by Dalenius and Reiss. Privacy in Statistical Databases, Proceedings: Annals of the New York Academy of Sciences 3050:14-29.

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

Reiter JP. 2004. New Approaches to Data Dissemination: A Glimpse into the Future. Chance 17(3):11-15.

Trottini M, Feinberg SE. 2002. Modeling User Uncertainty for Disclosure Risk and Data Utility. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 10(5):511-27. October, 2002.

Zayatz, L. 2007. Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update. Journal of Official Statistics 23(2):253-65.

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 **Appendix 7**. 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

The codebook documentation provides detailed information on each variable included on the 2006-2010 NSFG public use files. Each variable's codebook page contains all the basic information that was included in codebooks for Cycles 1-6 of the NSFG, with a few enhancements:

- Variable name
- Variable type (ENHANCED for 2006-2010)
- Column locations
- Question number and question text
- Universe statement
- Response categories and unweighted frequencies
- Link to recode specifications
- Notes, where applicable (ENHANCED for 2006-2010)

Each of these codebook elements is described further below.

The codebooks for each of the 3 public use data files for the 2006-2010 NSFG data release are available in 2 ways:

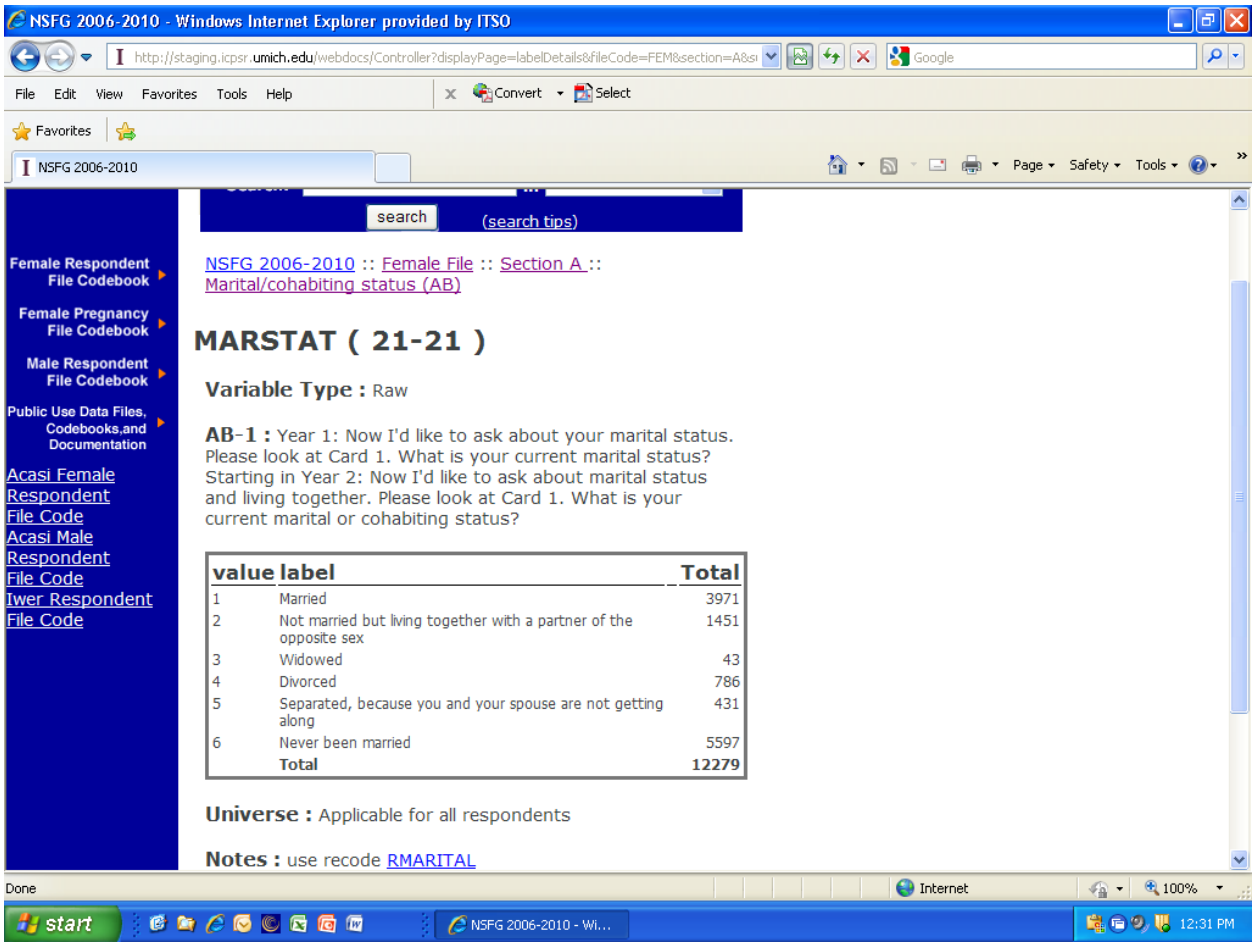
- **Online, interactive, searchable format with hyperlinks**: Web-based documentation, or "Webdoc," accessible through a link on the NSFG webpage
- **Stand-alone documents**: PDF format for downloading, saving, viewing, or printing from the NSFG webpage

Webdoc:

Webdoc is a Web-based tool that permits online, interactive access to the NSFG codebook documentation for all 3 public use files (female, pregnancy, and male). It 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.

A link to Webdoc is available through a link on the NSFG web site (www.cdc.gov/nchs/nsfg.htm). See below for an example page from Webdoc. This is the page displaying the detailed codebook information for the raw variable (and question) AB-1 MARSTAT. The specific elements of the codebook are described further below.



Codebooks for Cycle 6 (2002) NSFG are also available through Webdoc. The design and functioning of Webdoc for the 2006-2010 NSFG is essentially the same as that for Cycle 6 (2002), though some improvements have been made to facilitate searching for and printing the variables you need. For example, within the 2006-2010 Webdoc, you now have the option to generate an HTML file for specific sections or subsections of the codebook using the hyperlink at the bottom of the screen that says "Create codebook for <Section/Subsection>." The resulting file can be saved in PDF or other standard format, printed, or simply viewed on the screen.

PDF format:

In addition to Webdoc, the NSFG webpage provides access to the codebook documentation as downloadable, self-contained PDF documents, containing no active links to outside documents or sites. This may be useful when internet access is not possible or if you

wish to print specific sections of variables from the codebook. The codebook PDF files are available as 1 file per section of the questionnaire (e.g., female Section A, male Section C, etc.) except for Female Section E which is divided into sub-sections.

As noted above under "Webdoc," you also have the option to generate a file (HTML) for specific sections or subsections of the codebook within the interactive, online codebook tool. 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 Sterilization Operations (DA)* a screen with just the variables in the DA series from Female Section D is created. You can save this HTML file in PDF or other standard format, print it, or view it on the screen.

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 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 enter all that applied. The pregnancy file includes variables for all 4 mentions in PAYBIRTH1-4, 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 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 **recode** variable is a constructed variable created after the data is collected, from one or more raw variables, and has missing values imputed.
- An **intermediate** 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 during the interview, nor is it defined solely as a post-processing variable. Intermediate variables are defined in the process of constructing a recode that will have its missing values imputed.

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 2006-2010, including statements to assign appropriate variable and value labels.

Question Text:

Question text is either the actual question wording for a raw, asked question or the description of a computed, recode, or intermediate variable. The wording of the survey questions 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)**, is displayed, except that we have included the "Flow Check" number 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. Unless otherwise indicated, the Year 1 CRQs can be consulted to see these flow checks. 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.

Question text also includes any wording changes that were made during the course of 2006-2010 NSFG fieldwork. The year that the change took effect is included in the question text. *(See the Webdoc screen capture above for an example of how a Year 2 wording change is represented for AB-1 MARSTAT.)*

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 was not skipped by any respondents 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.)* **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.

In the 2006-2010 NSFG data files, questions or variables may be skipped or blank/sysmis for one of two reasons: 1) because the question or variable was not relevant or "applicable" for the respondent or 2) because the question was not included in the instrument when the respondent completed the interview.

1. Question or variable blank because it was not applicable at all: If a question or variable was not applicable to a particular respondent, the questionnaire program skipped to the next applicable question.

2. Question or variable blank because it was not applicable in this year of interviewing: Universe statements also indicate two types of changes to the questionnaire instrument during the course of interviewing. One type is a question addition. If a question was added to the instrument, only respondents whose interviews were completed after the addition will have data for the question. This is indicated in the universe statement by the description "Added in Year X," and the variable "QuestYear" is referenced. QuestYear indicates the "version year" of the questionnaire instrument, which roughly corresponds to calendar dates as follows:

   Year 1 = June 2006 through June 2007
   Year 2 = July 2007 to June 2008
   Year 3 = July 2008 to June 2009
   Year 4 = July 2009 to June 2010

   There were no question changes in year 4 and respondents in year 4 received the same questionnaire as year 3.The other type of questionnaire change indicated in the universe statement is a change in routing, or a change in the skip pattern involving a question. In this case, both or all universe statements are presented and the years that the change(s) took effect are included in the statement, along with a reference to QuestYear.

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. In the codebook files provided in PDF format on the NSFG webpage, 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.

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 with descriptive value labels and unweighted frequencies (or counts of cases). 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. The original values of these variables are intact in the file unless otherwise indicated. 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.

For a small percentage of variables, one or more response categories changed over the course of 2006-2010, either by having changes to the wording of one or more categories, or by having one or more categories added. If wording changed, both wording versions are presented and the year of the change is included in the response category text. If a response category was added, the wording for that category indicates the questionnaire year when it was not included and the year(s) that it was included, along with the text of the new category.

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 are not active in the 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 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 information the user should know about when using the variable.

DESCRIPTION OF QUESTIONNAIRES

The questionnaires for each year of the 2006-2010 NSFG are available in 2 formats:

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

Both of these formats are available as PDF files on the NSFG webpage. 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.

NSFG_2006-2010_Y1_Female_CapiLite.pdf
NSFG_2006-2010_Y2_Female_CapiLite.pdf
NSFG_2006-2010_Y3_Female_CapiLite.pdf
NSFG_2006-2010_Y1_Male_CapiLite.pdf
NSFG_2006-2010_Y2_Male_CapiLite.pdf
NSFG_2006-2010_Y3_Male_CapiLite.pdf

CAPI Reference Questionnaire (CRQ) Format

The CRQ shows all the detailed specifications necessary to program the NSFG questionnaires. As with the CAPI-Lite documents, there are CRQs for each survey year, however each section of the questionnaire is provided as a separate PDF file.

- 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 instructions for interviewers that accompanied many of the questions.

Because the CRQ is longer than the CAPI-lite for any questionnaire section, each section of the questionnaire for each year is provided as a separate PDF file on the NSFG webpage. For example, **NSFG_2006-10_Y1_FemaleA_CRQ.pdf** is the CRQ for female Section A for QuestYear=1, and **NSFG_2006-10_Y2_MaleF_CRQ.pdf** is the CRQ for male Section F for QuestYear=2. Please note that Female Section E is divided and has PDF files for each sub-section.

One goal of moving towards continuous interviewing for the NSFG was to enable changes to be made to the questionnaires whenever necessary, for example to correct programming errors or to add or revise questions. However, all substantive changes were implemented at the start of each data collection year of interviewing, particularly at the start of Year 1, with a much smaller set of changes at the start of Year 2 and Year 3. No changes were made at the start at Year 4. With the public use file release for 2006-2010, the Year 1 – Year 3 male and female CRQs have been posted for downloading, viewing, and printing from the NSFG website.

As noted above, there were no changes made to the CRQ for Year 4 – the Year 3 CRQs are applicable for Year 4 interviews. To represent changes within one CRQ would have made them too complex and difficult to follow, so each year is provided as it was fielded, as a "snapshot" of the questionnaire instrument during that particular period. Please consult **Appendix 5** for a summary of questionnaire changes that were implemented between Cycle 6 (2002) NSFG and Year 1 of the 2006-2010 NSFG, as well as between years 1 and 2 and years 2 and 3.

USER SUPPORT

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

By mail: NSFG Team
Reproductive Statistics Branch
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782

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

ACKNOWLEDGEMENTS

An adequate acknowledgement of all those who made significant contributions to the design, conduct, and production of this data file would require many pages. This brief acknowledgement will only name some of those who made important contributions.

The interviewing, data processing, and data file production for the 2006-2010 National Survey of Family Growth (NSFG) were conducted by the University of Michigan's Institute for Social Research (ISR), under a contract with the National Center for Health Statistics (NCHS).

At ISR, **Robert M. Groves** was the Project Director until April 2009, after which **James M. Lepkowski** assumed the role. **William G. Axinn** was the Deputy Project Director, and **Nicole Kirgis** served as Field Director and Director of Operations. James Lepkowski was the Chief Mathematical Statistician, until becoming Project Director, at which time **James Wagner** became Chief Mathematical Statistician. **Shonda Kruger-Ndiaye** played a key role in the development, testing, and evaluation of the survey instruments in the interviewer training and data collection phases of the project. The ISR data file and documentation production team was led by **Peter Granda**, Director of Data Processing, and included programmers **Michael Shove** and **Jason Langfahl.** The imputation programmers were **John Van Hoewyk** and **Brady West**.

At NCHS, the NSFG team, responsible for all aspects of NSFG survey design through data file and documentation production, was comprised of **William D. Mosher** (NSFG Project Officer), **Joyce C. Abma** (alternate Project Officer), **Anjani Chandra**, **Gladys M. Martinez,** and **Jo Jones. Brittany McGill** (until June 2008), **Paula Goodwin** (until April 2009), **Casey Copen** (since September 2008), **Veena Billioux** (since March 2010), **Kimberly Daniels** (since September 2010), and **Jonathan Vespa** (since September 2010) also made contributions to data file evaluation and documentation. Consultation on survey design, variance estimation and other statistical matters was provided by **Karen E. Davis** of the NCHS Office of Research and Methodology.

The 2006-2010 NSFG was jointly planned and funded by the following agencies of the U.S. Department of Health and Human Services:

- 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),
- Office of the Assistant Secretary for Planning and Evaluation (OASPE),
- CDC's Division of Reproductive Health (DRH),
- CDC's Division of HIV/AIDS Prevention (DHAP)
- CDC's Division of STD Prevention (DSTDP)
- Children's Bureau of the Administration for Children and Families (ACF);

NCHS is grateful for their assistance.

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

Sampling Errors in the 2006-2010 NSFG Examples:

The increases in sample size between the 2002 NSFG (n = 12,571) and the 2006-10 NSFG (n = 22,682) did not always result in a proportionate reduction in the variances of the statistics from the NSFG sample. ***In general,*** standard errors for comparable statistics with weighted denominators of the same size:

- Increased slightly for white women (example 1)
- Decreased significantly for black and Hispanic women (example 1)
- Decreased in some age groups but increased in others for women (example 2)
- Decreased significantly for men in all age and race groups (examples 3 & 4)
- Declined significantly for male teenagers (example 6)
- Declined slightly for female teenagers (example 5).

NOTE:
There are other ways to show the effects of design differences on sampling errors, and this Note has used just one approach–these comparisons used actual data, and therefore the denominators and percentages are not exactly the same in the 2002 and 2006-10 surveys. Another approach is to use artificial examples so that you can use exactly the same weighted denominators and exactly the same percentages in the two surveys; for example, 30 million women and exactly 20 percent in each survey. An example of this approach is in Advance Data from Vital and Health Statistics of the National Center for Health Statistics, #182, March 20, 1990, "Contraceptive Use in the United States, 1973-1988," tables II and III, page 8.

Example 1: Female File: Percent of women 15-44 who were childless at the date of interview, by race and Hispanic Origin

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All				
2002	7,643	61,561	41.6	(0.95)
2006-2010	12,279	61,755	44.4	(1.09)
Hispanic				
2002	1,589	9,107	32.4	(1.43)
2006-2010	2,723	10,474	35.5	(1.45)
Non-Hispanic White alone				
2002	4,038	39,498	44.2	(1.27)
2006-2010	6,156	37,384	47.7	(1.43)

Non-Hispanic Black alone				
2002	1,471	8,250	36.8	(1.62)
2006-2010	2,412	8,451	38.3	(1.67)
Non-Hispanic other (includes multiple-race respondents)				
2002	545	4,706	46.6	(3.00)
2006-2010	988	5,446	48.4	(3.16)

Example 1 uses all women in the samples as the denominator. It shows the percentage of all women who have not had any births as of the date of interview.

The percentage childless in 2006-2010 is 44.4%, and the standard error was 1.09—slightly larger in 2006-2010 than the 0.95 in 2002. For Hispanic women and black women, the standard errors are about the same size in 2002 and in 2006-2010. For white women, the difference is 1.43 vs. 1.27, or 0.16 percentage points, about a 13% increase in the standard error between 2002 and 2006-2010. This increase is a result of the data collection strategy to cope with a severely limited budget and still obtain large samples.

Example 2: Female File: Percent of women 15-44 who intend to have a(nother) birth, by age

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All				
2002	7,643	61,561	45.9%	0.89
2006-2010	12,279	61,755	48.8%	1.00
15-19				
2002	1,150	9,834	86.5%	1.28
2006-2010	2,284	10,478	89.1%	1.03
20-24				
2002	1,363	9,840	79.4%	1.50
2006-2010	2,098	10,365	79.5%	1.39
25-29				
2002	1,296	9,249	59.3%	1.79
2006-2010	2,366	10,535	62.3%	1.77
30-34				
2002	1,355	10,272	39.5%	1.56
2006-2010	2,047	9,188	38.4%	1.69
35-39				
2002	1,270	10,853	17.2%	1.30
2006-2010	1,798	10,538	17.7%	1.15
40-44				
2002	1,209	11,512	4.5%	0.74
2006-2010	1,686	10,652	5.6%	0.77

The denominator of the statistics in **example 2** is all women, but this time, by age. The standard errors in 2006-2010 for the age groups 15-19, 20-24, and 35-39 are smaller than in 2002. For example, at age 15-19, the standard error was reduced from 1.28 in 2002 to 1.03 for the same statistic—a nearly 20% drop. At age 20-24, the standard error declined from 1.50 to 1.39, a 7% decrease. At age 25-29 and 40-44, the standard errors were about the same in the 2 survey years. For the total (age 15-44), and for one age group, the standard errors were slightly higher in 2006-10 than in 2002.

Example 3: Male File: Percent of men 15-44 who intend to have a(nother) birth, by age

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All				
2002	4,928	61,147	55.4%	1.22
2006-2010	10,403	62,128	58.9%	1.00
15-19				
2002	1,121	10,208	89.5%	1.23
2006-2010	2,378	10,816	92.1%	0.79
20-24				
2002	938	9,883	85.0%	1.50
2006-2010	1,733	10,394	86.7%	1.37
25-29				
2002	708	9,226	71.8%	2.32
2006-2010	1,807	10,758	73.7%	1.64
30-34				
2002	724	10,138	47.6%	2.59
2006-2010	1,555	9,228	51.8%	2.03
35-39				
2002	746	10,557	29.1%	2.25
2006-2010	1,500	10,405	31.3%	1.98
40-44				
2002	691	11,135	16.5%	1.68
2006-2010	1,430	10,526	15.7%	1.37

The statistics shown in **Example 3** are based on all men, by age.
For men, the standard errors in 2006-2010 were substantially smaller than in 2002 in every age group. For example, at age 30-34, the standard error of the percentage who intended to father another child was 2.59 percentage points, and in 2006-2010, it was 2.03, about 22 percent smaller.

Example 4: MALE file: Percent of males 15-44 who have fathered 2 or more children

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All				
2002	4,928	61,147	29.6	(1.20)
2006-2010	10,403	62,128	29.0	(0.92)
Hispanic				
2002	1,123	10,188	38.2	(2.22)
2006-2010	2,409	11,847	36.7	(1.46)
Non-Hispanic White alone				
2002	2,526	38,738	26.8	(1.53)
2006-2010	5,275	37,283	26.2	(1.22)
Non-Hispanic Black alone				
2002	884	6,940	31.6	(1.91)
2006-2010	1,752	7,341	30.8	(1.52)
Non-Hispanic other (includes multiple-race respondents)				
2002	395	5,280	30.9	(3.55)
2006-2010	967	5,657	29.2	(3.41)

Example 4: For men, standard errors in the 2006-10 sample are uniformly and significantly smaller than in 2002. For all men, the standard error of the percentage is (1.20 - .92, or .28, or) 23% smaller. For Hispanic men, the standard error in 2006-2010 is 34% smaller (1.46 vs. 2.22). For white men, the standard error decreased from 1.53 to 1.22, or by 20%.
For black men, the standard error dropped from 1.91 to 1.52—or by 20%--from 2002 to 2006-2010.

Example 5: Female Teenagers 15-19: Percent of females 15-19 who have ever had sexual intercourse

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All females 15-19				
2002	1,123	9,598	45.5	1.8
2006-2010	2,255	10,361	42.6	1.7
Hispanic				
2002	220	1,447	37.4	3.6
2006-2010	520	1,849	42.1	3.3
Non-Hispanic White alone				
2002	578	5,930	45.1	2.6
2006-2010	1,160	5,979	41.7	2.3

Non-Hispanic Black alone				
2002	227	1,407	56.8	3.2
2006-2010	423	1,566	45.1	2.9

For female teenagers (Example 5), the standard errors were slightly smaller in 2006-2010 than in 2002 in each comparison. For example, for white female teens, the standard error is 2.3 in 2006-10, compared with 2.6 in 2002.

Example 6: Male Teenagers 15-19: Percent of males 15-19 who have ever had sexual intercourse

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All males 15-19				
2002	1,112	10,139	45.7	2.1
2006-2010	2,371	10,766	41.8	1.6
Hispanic				
2002	231	1,603	54.8	4.3
2006-2010	587	2,000	46.0	3.0
Non-Hispanic White alone				
2002	599	6,286	40.5	2.6
2006-2010	1,117	6,187	37.2	2.2
Non-Hispanic Black alone				
2002	190	1,347	61.9	4.7
2006-2010	437	1,582	59.4	3.1

Among male teenagers (Example 6), the standard errors dropped in 2006-2010.

For example, the standard error in 2006-2010 for all races was 1.6 compared with 2.1 in 2002, a 24% decrease. Among Hispanic and black male teens, the decreases in sampling errors were large
(For Hispanics, from 4.3 to 3.0 for Hispanics, a 30% reduction; and 4.7 vs. 3.1 for blacks, a 34% decrease).

Example 7: Pregnancy file: Percent of single live births in the last 5 years that were breastfed at all, by race of mother (based on BFEEDWKS = 0-995 and HISPRACE2)

	Denominator Sample (n)	Weighted # In Thousands	Weighted Percent	Standard Error
All				
2002	2,770	20,647	67.5%	(1.69)
2006-2010	4,499	20,774	69.3	(1.72)
Hispanic				
2002	745	4,177	74.7	(2.39)
2006-2010	1,224	4,513	75.0	(1.84)
Non-Hispanic White alone				
2002	1,299	1,2069	69.4	(2.37)
2006-2010	1,896	1,1337	72.9	(2.33)
Non-Hispanic Black alone				
2002	540	2,771	47.8	(2.90)
2006-2010	1,046	3,197	46.0	(2.63)
Non-Hispanic other (includes multiple-race)				
2002	186	1,630	68.4	(6.24)
2006-2010	333	1,728	72.5	(4.97)

Example 7 is from the pregnancy file, as reported by women in the NSFG sample. The standard error for the percentage of babies breast-fed is 1.72 in 2006-2010, compared with 1.69 in 2002—essentially the same. For babies born to Hispanic mothers, the standard error is 1.84 in 2006-2010, compared with 2.39 in 2002---a 23% decrease in the standard error. For babies born to non-Hispanic white women, the decrease was only .04, from 2.37 to 2.33, essentially no change. For births to black women, the decrease is from 2.90 to 2.63—a 9% decrease. So the sampling errors for percentages of births to Hispanic, black, and other women decreased, but there was no change in sampling errors for births to white women, or for all races.

ABORTION UNDER-REPORTING AMONG FEMALES IN THE NSFG

Abortions have always been under-reported in the National Survey of Family Growth (NSFG) and virtually all other demographic surveys. This has been determined by comparing NSFG weighted estimates of abortions with external data from abortion providers.

Numerator = Weighted number of abortions reported by NSFG respondents in a recent period such as 1996-2000 (2002 NSFG)

Denominator = Number of abortions reported in those same calendar years, based on data reported to CDC's Division of Reproductive Health or surveys of abortion providers conducted by the Guttmacher Institute.

Using this simple comparison, the estimated percentage of abortions reported by women 15-44, according to NSFG survey year, is shown below.

Cycle & survey year	Percent reported in the NSFG	
Cycle 2 (1976)	45%	
Cycle 3 (1982)	48%	
Cycle 4 (1988)	35%	
Cycle 5 (1995)	45%	(interview)
Cycle 5 (1995)	59%	(Self-administered questionnaire)
Cycle 6 (2002)	**47%**	**(for abortions in 1997-2001)**
2006-2008 NSFG	**35%**	**(for abortions in 2000-2005)**

(For further information on NSFG Cycle 6, see R.K Jones & K.Kost, "Underreporting of Induced and Spontaneous Abortion in the United States: An Analysis of the 2002 National Survey of Family Growth. *Studies in Family Planning* 38(3): 187-197. September 2007. Under-reporting based on the ACASI portion of the 2006-2008 NSFG has not yet been analyzed.)

CONCLUSION:
As in previous cycles, the NSFG staff advises NSFG data users that, generally speaking, NSFG data on abortion should **not** be used for substantive research. 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.

The study of the determinants and consequences of abortion is problematic and is, generally speaking, not advised.

BIRTH REPORTING AMONG FEMALES IN THE NSFG

The table below shows a comparison of NSFG-based estimates of births occurring in 2000-2006, relative to births registered through the vital records system in the United States for those same years. These figures are based on births reported by all female respondents in the 2006-2010 NSFG. To make the NSFG estimates comparable to vital records, we exclude any births reported by non-U.S.-born NSFG respondents that occurred before they came to the U.S. to stay. We also compare to vital records for births to women 44 or younger.

For the total number of births, as well as the numbers shown by race and Hispanic origin, mother's age at delivery, and birth order, the NSFG estimates do not differ significantly from vital records. That is, the ratios of NSFG to vital records do not differ significantly from 1, and the numbers based on vital records fall within the 95% confidence intervals for all of the NSFG estimates below.

	Number (in millions) of births from NSFG	95% Confidence Interval Lower Limit	95% Confidence Interval Upper Limit	Vital records	Ratio of NSFG/ Vital records
Total for 2000-2006	29.256	26.719	31.793	28.656	1.02
Race and Hispanic origin					
Hispanic or Latina	5.858	4.217	7.500	6.389	0.92
Not Hispanic or Latina:					
White	17.119	15.143	19.095	16.168	1.06
Black or African American	4.591	3.652	5.529	4.125	1.11
Mother's age at delivery					
15-19	3.268	2.712	3.825	3.020	1.08
20-24	7.503	6.580	8.425	7.249	1.04
25-29	7.767	6.898	8.637	7.711	1.01
30-44	10.653	9.554	11.752	10.641	1.00
Birth order					
1st	11.759	10.660	12.857	11.402	1.03
2nd	9.600	8.749	10.451	9.239	1,04
3rd or higher	7.897	6.855	8.940	8.015	0.99

Note: Given the larger sampling errors associated with individual-year birth estimates, we recommend users combine years when analyzing births reported in the NSFG.

References for the vital records data are the following National Vital Statistics Reports (NVSR) published annually by NCHS:

Martin JA et al. Births: Final data for 2006. NVSR 57(7). Hyattsville, MD: NCHS. 2009.
Martin JA et al. Births: Final data for 2005. NVSR 56(6). Hyattsville, MD: NCHS. 2007.
Martin JA et al. Births: Final data for 2004. NVSR 55(1). Hyattsville, MD: NCHS. 2006.
Martin JA et al. Births: Final data for 2003. NVSR 52(10). Hyattsville, MD: NCHS. 2005.
Martin JA et al. Births: Final data for 2002. NVSR 52(10). Hyattsville, MD: NCHS. 2003.
Martin JA et al. Births: Final data for 2001. NVSR 51(2). Hyattsville, MD: NCHS. 2002.
Martin JA et al. Births: Final data for 2000. NVSR 50(5). Hyattsville, MD: NCHS. 2002.

SEX EDUCATION ITEMS IN THE NSFG

The following table presents, for 2006-2010 and earlier NSFG Cycles, the topics included within the series on formal instruction on sex education and talking with parents about sex education topics. The National Survey of Adolescent Males (NSAM) is included because it provides data for male teens that are parallel to data provided for female teens in the 1995 NSFG.

Sex education topics included in the NSFG, Cycles 5, 6, and 7, and the 1995 NSAM							
	FEMALES				MALES		
	1995	2002	2006-2010		1995 (NSAM)	2002	2006-2010
Universe: ages =	15-44	15-19	15-24		15-19	15-19	15-24
	Talked to parent about[1]				Talked to parent about		
[3]How to say no to sex		x	X			x	x
[3]Methods of birth control	x	x	X		x	x	x
Where to get birth control		x	X			x	x
[3]How to prevent HIV/AIDS			X				x
How to use a condom		x	X			x	x
[3]Sexually transmitted diseases	x	x	x			x	x
How pregnancy occurs	x						
Other sexually transmitted diseases like herpes or gonorrhea (STD)					x		
AIDS					x		
What would happen if you got a girl pregnant					x		
	Formal Instruction[2]				Formal instruction		
[3]How to say no to sex[4]	x	x	x		x	x	x
[3]Methods of birth control	x	x	x		x	x	x
[3]Sexually transmitted diseases	x		x				x
[3]How to prevent HIV/AIDS			x				x
How to prevent AIDS using safe sex practices	x				x		
Other sexually transmitted diseases like herpes or gonorrhea					x		
AIDS					x		
Where to get birth control					x		
How to put on a condom/ how to use a condom					x		

[1]Before you were 18 years old, which, if any of the topics did you ever talk with a parent or guardian about?
[2]Before you were 18, did you ever have any formal instruction at school, church, a community center or some other place about….
[3]Healthy People 2010 and 2020 objective
[4]1995 wording: "Abstinence or how to say no to sex?"

RELIGION DATA IN THE NSFG

All NSFG respondents were asked "In what religion were you raised, if any?" and "What is your current religion, if any?" For both these questions, they were shown a card listing the 10 religious groups with the largest memberships in the United States (RELNOW and RELRSD). If their religion was not among those listed, the respondents were shown a 2nd card with 17 more religions (RELNOW1 and RELRSD1), plus an "Other-specify" option. A small number of cases provided "Other-specify" responses, and these verbatim responses (OTHRLNOW and OTHRLRSD) were reviewed to determine if any could be "back-coded" into one of the existing detailed categories or if a new category should be created. In the 2006-2010 file, there are too few cases to do analysis with the detailed religion categories. This may change as more cases are collected.

The coding of the other-specify religion verbatim responses into existing and new categories of RELRSD/RELRSD1 (religion in which respondent was raised) and RELNOW/RELNOW1 (current religion) followed the same logic for males as for females. Although the variable names and question routing are the same in both instruments, the question numbers are different because these questions were asked in different sections – Section I for females and Section J for males. For ease of reference, the table below presents a crosswalk of the variable names with the question numbers for males and females. Besides the religious affiliation variables described above, the table also lists FUNDAM1, a question asking whether the respondent believes certain religious labels apply to him or her.

Variable Name	Female Question Number	Male Question Number
RELRSD	IC-1	JB-1
RELRSD1	IC-2	JB-2
OTHRLRSD	IC-3	JB-3
RELNOW	IC-5	JB-5
RELNOW1	IC-6	JB-6
OTHRLNOW	IC-7	JB-7
FUNDAM1	IC-8	JB-8

RELCURR and RELRAISD: "Current religion" and "Religion raised"

Respondents were allowed to enter verbatim responses for the religion in which they were raised (OTHRLRSD) and the religion they practiced at the time of the interview (OTHRLNOW) if they could not find the response category for their religion on the show cards for RELRSD/RELRSD1 and RELNOW/RELNOW1.

The "Other-specify" verbatim responses for OTHRLNOW were coded into existing and new categories of RELNOW and RELNOW1. Once all verbatim responses were assigned to an existing or new category, RELNOW and RELNOW1 were used to create **RELCURR** for the public use file. The full detail of RELNOW and RELNOW1 is available through the NCHS Research Data Center (RDC).

Similarly, after "Other-specify" responses for OTHRLRSD were coded into existing and new categories of RELRSD and RELRSD1, these variables were used to create the variable **RELRAISD** for the public use file. The full detail of RELRSD and RELRSD1 is available

through the RDC.

RELCURR and RELRAISD are both variables constructed in post-processing for inclusion on the public use file, and do not have imputed values. RELCURR also serves as an intermediate variable in the construction of the RELIGION recode (see also section below on **VARIABLES MODIFIED OR SUPPRESSED FOR PUBLIC USE** and **Appendix 3a (female)** and **3c (male) recode specifications**).

Details of the construction of RELCURR from raw variables RELNOW and RELNOW1 follow. RELRAISD was created in the same manner as RELCURR using raw variables RELRSD and RELRSD1.

-- RELCURR=1 if RELNOW = None (1) or RELNOW1 = No particular faith (90)

-- RELCURR=2 if RELNOW = Catholic (2)

-- RELCURR=3 if RELNOW = Southern Baptist (4) or Baptist (5)

-- RELCURR=4 if RELNOW = Methodist or African Methodist (6), Lutheran (7), Presbyterian (8), or Episcopal or Anglican (9)

-- RELCURR=5 if RELNOW1 = Assemblies of God (12), Church of Nazarene (13), The Church of God (14), The Church of God (Cleveland, TN) (15), The Church of God in Christ (16), 7th Day Adventist (17), United Pentecostal Church (18), Pentecostal Assemblies (19), or Fundamental Protestant Bodies, Pentecostal (30)

-- RELCURR=6 if RELNOW1 = Christian, another denomination not listed (21)

-- RELCURR=7 if RELNOW1 = Christian, no specific denomination (22)

-- RELCURR=8 if RELNOW = Jewish (3) or Church of Jesus Christ of Latter Day Saints, (LDS/Mormon) (10), or

if RELNOW1 = Jehovah's Witness (20), Unitarian-Universalist (23), Greek Orthodox (24), Other Orthodox (25), Muslim (26), Buddhist (27), Hindu (28), Native American religions (31), Taoic religions (32), Neopagan religions (33), or Other, not shown separately (95)

-- RELCURR=9 if RELNOW = Refused (98)

-- RELCURR=10 if RELNOW = Don't know (99)

Code categories:

1 = No religion
2 = Catholic
3 = Baptist/Southern Baptist
4 = Methodist, Lutheran, Presbyterian, Episcopal
5 = Fundamentalist Protestant
6 = Other Protestant denomination

7 = Protestant - No specific denomination
8 = Other religion
9 = Refused
10 = Don't know

RELTRAD: "Current religious affiliation by Protestant denomination"
(Please see above table for question numbers.)

The variable RELTRAD categorizes Protestant respondents into three groups depending on the respondent's Protestant denomination (RELNOW and RELNOW1), his or her fundamentalist ideology (FUNDAM1), and race (recode RACE). RELTRAD was constructed in post-processing and missing values were not imputed. (In specifications below, "ne" means "not equal to.") We have included RELTRAD on the 2006-2010 NSFG public use file because many scholars have found it a useful categorization of religious affiliation, "grounded in the historical development of American religious traditions" *(Steensland B, Park JZ, Regnerus MD, Robinson LD, Wilcox WB, Woodberry RD. The Measure of American Religion: Toward Improving the State of the Art. Social Forces 79(1):291-318. 2000)*.

-- RELTRAD=1 if RELNOW = (Southern Baptist (4) or Baptist (5)) **and** RACE ne black (1), or

if RELNOW1 = Assemblies of God (12), Church of Nazarene (13), The Church of God (14), The Church of God (Cleveland, TN) (15), 7th Day Adventist (17), United Pentecostal Church (18), Pentecostal Assemblies (19), or Fundamental Protestant Bodies, Pentecostal (30), or

if RELNOW1 = (Christian, another denomination not listed (21) or Christian, no specific denomination (22)), **and** RACE ne black (ne 1), **and** FUNDAM1 = a born again Christian, a charismatic, an evangelical, or a fundamentalist (1, 2, 3, 4)

-- RELTRAD=2 if RELNOW = Methodist or African Methodist (6) **and** RACE ne black (ne 1), or

if RELNOW = Lutheran (7), Presbyterian (8), Episcopal or Anglican (9), or

RELNOW1 = (Christian, another denomination not listed (21) or Christian, no specific denomination (22)), **and** RACE ne black (ne 1), **and** FUNDAM1 = none of the above, dk, rf (5, 8, 9),

-- RELTRAD=3 if RELNOW = (Southern Baptist (4), Baptist (5), or Methodist or African Methodist (6)) **and** RACE = black (1), or

if RELNOW1 = The Church of God in Christ (16), or

RELNOW1 = (Christian, another denomination not listed (21) or Christian, no specific denomination (22)) **and** RACE = black (1),

-- RELTRAD=4 if RELNOW = Catholic (2)

-- RELTRAD=5 if RELNOW = Jewish (3) or Church of Jesus Christ of Latter Day Saints, (LDS/Mormon) (10), or

if RELNOW1 = Jehovah's Witness (20); Unitarian-Universalist (23); Greek Orthodox (24); Other Orthodox (25); Muslim (26); Buddhist (27); Hindu (28); Native American religions (31); Taoic religions (32); Neopagan religions (33); or Other,-not shown separately (95)

-- RELTRAD=6 if RELNOW = None (1) or RELNOW1 = No particular faith (90)

-- RELTRAD=8 if RELNOW = Refused (98)

-- RELTRAD=9 if RELNOW = Don't know (99)

Code categories:

1 = Evangelical Protestant
2 = Mainline Protestant
3 = Black Protestant
4 = Catholic
5 = Other religion
6 = No religious affiliation
8 = Refused
9 = Don't know

VARIABLES MODIFIED OR SUPPRESSED FOR PUBLIC USE

As noted in Part 1 of the User's Guide, several variables collected in the 2006-2010 NSFG had to be modified or suppressed for the public use files to minimize the risk of disclosure. These variables are all enumerated in **Appendix 7**, and the full-detail, original versions of these analytic variables are available through the NCHS Research Data Center (RDC). Most of the suppressed variables have been included on the NSFG public use files in some form -- e.g., with fewer categories, with some values top- or bottom-coded, or with new variables to summarize information from several variables. In some cases, the public use variable and the RDC variable are named similarly, with the RDC variable distinguished by an "IN" prefix.

Please refer to the User's Guide section called "**Protections to Minimize Risk of Disclosure for Individual-Level Data**" for an overview of the measures taken to protect the confidentiality of NSFG respondents.

Below is a brief description of the variables in the 2006-2010 NSFG that were modified or suppressed for public use, to minimize the risk of disclosure. First, those variables found on both the female respondent and male respondent files are described, roughly in the order they occur in the questionnaires. Then the variables applicable only to the female or male files are described, also in section order. References are included to help you find further information in the questionnaires or recode specifications.

Variables Modified or Suppressed on Both Female and Male Files:

- **ROSCNT/INROSCNT** *(number of household members)* **–** The computed variable ROSCNT (female A-15 and male A-12) was topcoded at "8 or more household members" for public use. The full detail is retained in INROSCNT.

- **CSPBBHH/INCSPBBHH** *(number of R's biological children (aged 18 or younger) with current husband or cohabiting partner who live in the household)* – This Section A recode was topcoded at "3 or more joint biological children in the household" for public use. The full detail is retained in INCSPBBHH. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **CSPOKDHH/INCSPOKDHH** *(number of all other children (aged 18 or younger) in household living with R and her current husband or cohabiting partner)* – This Section A recode was topcoded at "1 or more children under 19 in household" for public use. The full detail is retained in INCSPOKDHH. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **NUMKDHH/ INNUMKDHH** *(number of biological/adopted/related/legal children under age 18 in household)* **–** This Section A recode was topcoded at "5 children or more" for public use. The full detail is retained in INNUMKDHH. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **NUMFMHH/ INNUMFMHH** *(number of family members in household)* **–**This Section A recode was topcoded at "7 or more family members" for public use. The full detail is

retained in INNUMFMHH. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **HISPGRP/ INHISPGRP** *(Hispanic subgroup)* **--** The categories for the raw variable AC-2 HISPGRP were collapsed for public use into: 1) Mexican; and 2) all other Hispanic groups. The full detail is retained in INHISPGRP.

- **Respondent's race** (AC-3 RRACE1-RRACE5, AC-4 RACEBEST, AC-5 OBSERVE) – NSFG respondents were allowed to report as many racial categories as may apply to them, and the maximum number reported was 4 (AC-3 RRACE1-5). In the 2006-2010 NSFG, 492 females (4.0%) and 492 of males (4.7%) reported more than 1 race. For those who reported more than 1 race, AC-4 RACEBEST asked them to identify the race that best describes them. For those who answered "don't know" or "refused" on the race questions, AC-5 OBSERVE indicates how the interviewer classified the respondent's race by observation. For the public use file, all of these race variables, along with variables for Hispanic origin, were used to create 2 recodes that describe race or Hispanic origin: (HISPANIC, RACE) as well as 2 recodes that describe Hispanic origin and race (HISPRACE, HISPRACE2). HISPRACE is the race and Hispanic origin recode that complies with the **1977** Office of Management and Budget (OMB) guidelines, in which multiple-race respondents are combined with single-race respondents. HISPRACE2 is the race and Hispanic origin recode that complies with the **1997** OMB guidelines, in which multiple-race respondents are separately classified. The NUMRACE intermediate variable included on the public use file indicates whether the respondent reported multiple races. See specifications in **Appendix 3a (female)** or **3c (male)** for more detail on NUMRACE and these 4 recodes. The original race variables are available through the NCHS RDC.

- **Household roster variables** (AD-2 USUALRES[x], AD-3 SEX[x], AD-4 AGE[x], AD-5 RELAR[x], AD-9 RELMAN [x] / RELWOM[x], computed BOTHBIOL and computed BOTHADOP (female A-24 and male A-19), –These raw and computed variables specify the usual residence of the respondent, as well as the gender, age, and relationship of each household member to the respondent and family structure. Several recodes were created from these raw household roster variables *(for example: HHPARTYP (Type of parental situation in household) and HHKIDTYP (Whether R has children (18 or younger), and whether bio/non-bio, living in household))*. See **Appendix 3a (female)** or **Appendix 3c (male),** Section A recodes, for more detail on these recodes.

- **HIGRADE/ INHIGRADE** *(current grade in school or highest grade/yr attended)* The categories for the raw variable HIGRADE (female AF-3 and male AE-3) were bottomcoded at "9th grade or less" for public use. The full detail is retained in INHIGRADE.

- **EDUCAT/ INEDUCAT** *(number of years of schooling)* – This Section A recode was bottomcoded at "9th grade or less" for public use. The full detail is retained in INEDUCAT. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **HIEDUC/ INHIEDUC** *(highest completed year of school or highest degree received***)** – This Section A recode was bottomcoded at "9th grade or less" for public use. The full

detail is retained in INHIEDUC. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **PARAGE14/ INPARAGE14** *(parental living situation at age 14)* – The categories for this Section A recode were collapsed for public use into: 1) respondent lived with both biological parents at age 14; 2) respondent lived with biological mother and step-father at age 14; and 3) respondent lived in any other parental situation or a non-parental situation at age 14. The full detail is retained in INPARAGE14. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **LVSIT14F/ INLVSIT14F** *(female parent/parent-figure at age 14)* – The categories for the raw variable LVSIT14F (female AG-3 and male AF-3) were collapsed for public use into: 1) biological mother; 2) other mother figure; 3) no mother figure. The full detail is retained in INLVSIT14F.

- **LVSIT14M/ INLVSIT14M** *(male parent/parent-figure at age 14)* – The categories for the raw variable LVSIT14M (female AG-4 and male AF-4) were collapsed for public use into: 1) biological father; 2) step-father; 3) no father figure; 4) other father figure. The full detail is retained in INLVSIT14M.

- **WTHPARNW/ INWTHPARNW** *(R is living with parents or parent figures)* – The categories for the computed variable WTHPARNW (female A-24 and male A-19), were collapsed for public use into: 1) both biological or adoptive parents; 2) other or no parental figures. The full detail is retained in INWTHPARNW.

- **WOMRASDU/ INWOMRASDU** *(woman who raised R during teens)* – The categories for the raw variable WOMRASDU (female AG-5 and male AF-5) were collapsed for public use into: 1) biological mother; 2) other mother figure; 3) no mother figure. The full detail is retained in INWOMRASDU.

- **MOMDEGRE/ INMOMDEGRE** *(highest level of education female parent (figure) completed)* – The categories for the raw variable MOMDEGRE (female AG-6 and male AF-6) were collapsed for public use into: 1) less than high school; 2) high school graduate or GED; 3) some college, including 2-year degrees; 4) bachelor's degree or higher. The full detail is retained in INMOMDEGRE.

- **EDUCMOM/ INEDUCMOM** *(mother's (or mother-figure's) education)*– The categories for this Section A recode were collapsed for public use into: 1) less than high school; 2) high school graduate or GED; 3) some college, including 2-year degrees; 4) bachelor's degree or higher; 5) no mother/mother-figure identified (coded as "95"). The full detail is retained in INEDUCMOM. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **MOMCHILD/ INMOMCHILD** *(number of children born to female parent (figure)* **–** The raw variable MOMCHILD (female AG-8 and male AF-8) was topcoded at "6 children or more" for public use. The full detail is retained in INMOMCHILD.

- **MOMFSTCH/ INMOMFSTCH** *(age of female parent (figure) at her first birth)* **–** The raw variable MOMFSTCH (female AG-9 and male AF-9) was collapsed for public

use into: 1) less than 18; 2) 18-19 years; 3) 20-24 years; 4) 25-29 years; 5) 30 years or older. The full detail is retained in INMOMFSTCH.

- **AGEMOMB1/ INAGEMOMB1** (*age of mother (or mother-figure) at first birth)* – The values of this Section A recode were collapsed for public use into: 1) less than 18; 2) 18-19 years; 3) 20-24 years; 4) 25-29 years; 5) 30 years or older. The full detail is retained in INAGEMOMB1. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **MANRASDU/ INMANRASDU** *(man who raised R during teens)* – The categories for the raw variable MANRASDU (female AG-11 and male AF-11) were collapsed for public use into: 1) biological father; 2) stepfather; 3) no father figure; 4) other father figure. The full detail is retained in INMANRASDU.

- **DADDEGRE/ INDADDEGRE** *(highest level of education male parent (figure) completed)* – The categories for the raw variable DADDEGRE (female AG-12 and male AF-12) were collapsed for public use into: 1) less than high school; 2) high school graduate or GED; 3) some college, including 2-year degrees; 4) bachelor's degree or higher. The full detail is retained in INDADDEGRE.

- **LIFPRTNR/ INLIFPRTNR** *(number of opposite-sex sexual partners in lifetime)* – This recode was topcoded at "50 or more opposite-sex partners in lifetime" for public use. The full detail is retained in INLIFPRTNR. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **Race and Hispanic origin of spouse or partner –** Because the OMB requires all federal surveys to collect multiple races whenever race is collected, the NSFG collected more detail than we can show on the public use file, given our sample sizes. For both male and female respondents, we created new, collapsed versions of the variables for Hispanic origin and race of their spouses and partners, so that the public use file shows the same categories of race and origin for the spouse or partner as it does for the respondent (see above for "Respondent's race"). These collapsed variables greatly reduce the number of small cells on these characteristics, and are analogous to RACE, NUMRACE, HISPRACE, and HISPRACE2 that describe respondents themselves.

  <u>For females</u>: Collapsed versions of 4 variables were created for
  (a) each reported husband (if any),
  (b) the first cohabiting partner (if applicable),
  (c) the current cohabiting partner (if applicable),
  (d) the first sexual partner ever (if not also a husband or cohabiting partner), and
  (e) up to 3 <u>current</u> sexual partners in the last 12 months (if applicable).

  <u>For males</u>: The same set of 4 collapsed public use variables was created from the more detailed information in the questionnaires. Since the marriage history information is collected differently, the variables for males were for
  (a) the current wife or cohabiting partner,
  (b) the first wife (if she isn't the current wife),
  (c) the first cohabiting partner (if she isn't the current one), and

(d) up to 3 current sexual partners in the last 12 months (if he isn't married or cohabiting).

	RACE-like variable	Multiple Race variable	HISPRACE-like variable	HISPRACE2-like variable
FEMALE FILE				
Husbands	HsbRACE1-5	HsbMult1-5	HsbHRACE1-5	HsbNRACE1-5
Current Cohabiting Partner	CurCohRACE	CurCohMult	CurCohHRACE	CurCohNRACE
1st Cohabiting Partner	Coh1RACE	Coh1Mult	Coh1HRACE	Coh1NRACE
1st Sexual Partner	fsexRACE	fsexMult	fsexHRACE	fsexNRACE
Recent Sexual Partners	p1yRACE1-3	p1yMult1-3	p1yHRACE1-3	p1yNRACE1-3
MALE FILE				
Current Wife or Cohabiting Partner	cwpRACE	cwpMult	cwpHRACE	cwpNRACE
1st Wife	Wif1RACE	Wif1Mult	Wif1HRACE	Wif1NRACE
1st Cohabiting Partner	Coh1RACE	Coh1Mult	Coh1HRACE	Coh1NRACE
Recent Sexual Partners	p1yRACE1-3	p1yMult1-3	p1yHRACE1-3	p1yNRACE1-3

All of the original, full-detail race and Hispanic origin variables that were used to define the variables for each of the listed spouses and partners above are available through the RDC. These include (in order corresponding to the table above):

- Female variables: CB-8 HISPHX, CB-9 RACEHX, CB-10 BSTRACHX. CC-8 CPHISP, CC-9 CPRACE, CC-10 CPBESTR, CD-9 HISPCX, CD-10 RACECX, CD-11 BSTRACCX, CG-7c FPHISP, CG-7d FPRACE, CG-7e FPRACEB, CI-14 P1YHISP, CI-15 P1YRACEB, CI-16 P1YRACEX
- Male variables: CB-3 CWPHISP, CB-4 CWPRACE, CB-5 CWPRACEB, EC-3 FWPHISP, EC-4 FWPRACE, EC-5 FWPRACEB, DD-15 PXHISP, DD-16 PXRACE, DD-17 PXBEST

- **YRSTRUS/CMSTRUS** *(year/century month R came to the United States to stay*) **–** The raw variables for the month and year the respondent came to the United States to stay (Male variables: JA-8m STRUS_M and JA-8y STRUS_Y; female variables: IB-9m STRUS_M and IB-9y STRUS_Y) and corresponding century month variable computed in Flow Check J-2/I-5 (cmstrus) are not included on the public use file due to disclosure risk. A variable indicating only the year the respondent came to the U.S. to stay (**yrstrus**) was created for public use. The original, full-detail variables (STRUS_M/_Y and cmstrus) are available through the RDC.

- **PAYDU/INPAYDU** *(current living quarters owned/rented, etc)* **–**The categories for the raw variable PAYDU (female IB-10 and male JA-9) were collapsed for public use into: 1) Owned or being bought by you or someone in your household and 2) Rented. The full detail is retained in INPAYDU.

- **RELRAISD** *(religion in which respondent was raised)* – RELRAISD is a variable constructed for public use (in post-processing) from the raw religion variables RELRSD (fem IC-1/male JB-1) and RELRSD1 (fem IC-2/male JB-2). The coding and categories for RELCURR are shown above in **RELIGION DATA IN THE NSFG**. The original, full-detail variables RELRSD and RELRSD1 are available through the RDC.

- **RELCURR** *(current religious affiliation)* – RELCURR is an intermediate variable

created for public use from the raw religion variables RELNOW (fem IC-5/male JB-5) and RELNOW1 (fem IC-6/male JB-6). RELCURR is used to define the RELIGION recode (see **Appendix 3a (female)** and **3c (male)** for further information on the RELIGION recode). The coding and categories for RELCURR are shown above in **RELIGION DATA IN THE NSFG**. The original, full-detail variables RELNOW and RELNOW1 are available through the RDC.

- **DOLASTWKn/INDOLASTWKn** – Because respondents frequently engage in multiple activities, e.g. work and school, in a given week, they were shown a list of 9 activities and asked an "enter all that apply" question about what they had done in the previous week (DOLASTWK1-9 (female IE-1 and male JE-1). The previous week was defined as the week ending the Saturday before the interview. This created 9 distinct variables, each of which could contain any of these 9 responses:

  1) Working
  2) Not working at job due to temporary illness, vacation, strike, etc
  3) On maternity, paternity or family leave from job
  4) Unemployed, laid off, or looking for work
  5) Keeping house
  6) Taking care of family
  7) Going to school
  8) On permanent disability
  9) Something else

  No respondent in the 2006-2010 NSFG entered more than 6 activities in the previous week, so only DOLASTWK1-6 are on the file. The 9 possible categories were collapsed as follows for public use:

  If INDOLASTWKn = 1 then DOLASTWKn = 1
  If INDOLASTWKn = 2 or 3 then DOLASTWKn = 2
  If INDOLASTWKn = 4 then DOLASTWKn = 3
  If INDOLASTWKn = 5 or 6 then DOLASTWKn = 4
  If INDOLASTWKn = 7 then DOLASTWKn = 5
  If INDOLASTWKn = 8 or 9 then DOLASTWKn = 6

  The collapsed code categories are:

  1) Working
  2) Working – maternity/paternity leave or temp leave
  3) Not working, looking for work
  4) Keeping house or taking care of family
  5) In school
  6) Other

  The original, full-detail variables, named with an "IN" prefix, are available through the RDC.

- **LABORFOR/ INLABORFOR** *(labor force status)* -- The values of this recode (female Section I; male Section J) were collapsed for public use from 10 categories into 9 categories. The full detail is retained in INLABORFOR. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **SPLSTWKn/ INSPLSTWKn** – The raw variables SPLSTWKn (female IF-1 and male JF-1) indicate what activities were reported by each married or cohabiting respondent for his/her current spouse or partner in the week ending the Saturday before the interview.

Similar to DOLASTWK, there were 9 variables created to hold the "enter all that apply" responses, and the 9 possible responses were the same as for DOLASTWKn (shown above).

No respondent in the 2006-2010 NSFG entered more than 5 activities in the previous week for his/her spouse or cohabiting partner, so only SPLSTWK1-5 are on the file. The 9 possible categories were collapsed as follows for public use:

If INSPLSTWKn = 1 then SPLSTWKn = 1
If INSPLSTWKn = 2 or 3 then SPLSTWKn = 2
If INSPLSTWKn = 4 then SPLSTWKn = 3
If INSPLSTWKn = 5 or 6 then SPLSTWKn = 4
If INSPLSTWKn = 7, 8, or 9 then SPLSTWKn = 5

The collapsed code categories are:

1) Working
2) Working – paternity leave or temp leave
3) Not working, looking for work
4) Keeping house or taking care of family
5) All other

The original, full-detail variables, named with an "IN" prefix, are available through the RDC.

- **INCHES/RWEIGHT/BMI** *(self-reported height and weight and computed Body-Mass Index)* – In the Audio CASI portion of the 2006-2010 NSFG interview, respondents were asked their current height and weight (female JA-5 RHEIGHT_FT/_IN & JA-6 RWEIGHT; male RHEIGHT_FT/_IN & JA-6 RWEIGHT). In Cycle 6, similar variables were included on the ACASI file, but are being provided on the main public use files now. For respondent's height, a new variable called INCHES was defined using RHEIGHT_FT and RHEIGHT_IN values. Due to some extremely low and some extremely high values reported on height and weight, INCHES and RWEIGHT (weight in pounds) have been modified for public use, based on the 5^{th} and 95^{th} percentiles of their distributions, separately for males and females.
  - INCHES for females has been bottomcoded at "59 inches or less" and topcoded at "71 inches or more" (that is, 4'11" and 5'11").
  - INCHES for males has been bottomcoded at "63 inches or less" and topcoded at "77 inches or more."
  - RWEIGHT for non-pregnant female respondents has been bottom-and top-coded at 104 pounds and 275 pounds, respectively.
  - RWEIGHT for males has been bottom- and top-coded at 124 and 300 pounds, respectively.

  The original ININCHES and INRWEIGHT variables are available through the RDC. For most respondents (males 20-44 years of age and non-pregnant females 20-44 years of age), Body-Mass Index, or BMI, has also been computed and included on the public use files. BMI is a commonly used measure for assessing overweight or underweight of (non-pregnant) adults. It is defined by the following formula:

  ROUND [[(RWEIGHT) / (INCHES)**2] * 703]

  For cases with top- or bottom-coded values on INCHES and RWEIGHT, the modified values were used in the computation of BMI.

  BMI is not computed for pregnant women because their weights at the time of interview do not reflect their usual or "normal" weight. For teenage respondents (males and females 15-19 years of age), BMI computed by the above formula is also not considered an appropriate assessment tool for overweight or underweight. It is more

appropriate to assess weight for height, relative to standardized growth curves by age. For more information on the latest growth curves estimated by the Centers for Disease Control and Prevention, please visit http://www.cdc.gov/nchs/data/nhanes/databriefs/growthch.pdf.

- **POVERTY/ INPOVERTY** *(poverty level income)* – This recode was topcoded at "500 percent or more of poverty" for public use. The full detail is retained in INPOVERTY. See recode specifications in **Appendix 3a (female)** or **Appendix 3c (male)**.

- **PUBASTYP1/ INPUBASTYP1** – The raw variable PUBASTYP (female JI-5 and male KL-5) asked respondents whether they had received 4 different types of cash assistance:
  1) [State-specific Temporary Assistance for Needy Families (TANF)]/ welfare/ AFDC,
  2) General Assistance
  3) Emergency Assistance/ short-term cash assistance
  4) Some other program

  Because this was an "enter all that apply" variable, 4 mentions (variables) were allowed. Only one PUBASTYP1 variable is included on the public-use files and the categories were collapsed into the following 2 categories, based on any reporting of TANF/ welfare/ AFDC in any of the 4 variables:
  1) Public assist program, e.g. AFDC or ADC
  2) General/Emergency/Other assistance

  The full detail is retained in INPUBASTYP1.

Female-File-Only Variables Modified or Suppressed for Public Use:

- **NUMCHILD/INNUMCHILD** *(number of children in household < 13 yrs)* – The values for NUMCHILD computed in Flow Check A-15 were topcoded at "4 children or more" for public use. The full detail is retained in INNUMCHILD.

- **HHKIDS18/INHHKIDS18** *(number of children in household < 19 yrs)*-- The values for HHKIDS18 computed in Flow Check A-15 were topcoded at "5 children or more" for public use. The full detail is retained in INHHKIDS18.

- **NONBIOKIDS/INNONBIOKIDS** *(number of children in household < 19 yrs)* – The values for NONBIOKIDS computed in Flow Check A-15 were topcoded at "2 children or more" for public use. The full detail is retained in INNONBIOKIDS.

- **DAUGHT918/INDAUGHT918** *(number of daughters in household 9-18 yrs)* -- The values for DAUGHT918 computed in Flow Check A-15, starting in Year 2 of the 2006-2008 NSFG (July 2007), were topcoded at "2 children or more" for public use. The full detail is retained in INDAUGHT918.

- **HPLOCALE/ HPLOCATN** *(where husband/partner currently lives, if not on household roster)* -- The categories for the raw variable AD-8 HPLOCATN were collapsed for a new public use file variable HPLOCALE showing: 1) lives in the household; or 2) lives elsewhere. The full detail is retained in HPLOCATN.

- **CSPBSHH/INCSPBSHH** *(number of female R's biological children (aged 18 or*

younger) in household who are not the biological children of her current husband or cohabiting partner)– This Section A recode was topcoded at "1 or more children under 19 in household" for public use. The full detail is retained in INCSPBSHH. See recode specifications in **Appendix 3a**.

- **Numbers of male sexual partners in lifetime** (CH-1 LIFEPRT/ INLIFEPRT; CH-1b LIFEPRT_LO/ INLIFEPRT_LO; CH-1c LIFEPRT_HI/ INLIFEPRT_HI; computed LIFEPRTS/INLIFEPRTS from Flow Check C-83) – These raw and computed variables were all topcoded at "50 or more partners" for public use. The full detail is retained in the 4 "IN" variables.

- **Number of male sexual partners before 1st marriage (including husband)** (CH-2 PTSB4MAR/ INPTSB4MAR; CH-2b PTSB4MAR_LO/ INPTSB4MAR_LO; CH-2c PTSB4MAR_HI/ INPTSB4MAR_HI) – These 3 raw variables were topcoded at "50 or more partners for public use. The full detail is retained in the 3 "IN" variables.

Male-File-Only Variables Modified or Suppressed for Public Use:

- **WPLOCALE/ WPLOCATN** *(where R's current wife/partner lives*) -- The categories for the raw variable AD-8 WPLOCATN were collapsed for a new public use file variable WPLOCALE showing: 1) lives in the household; or 2) lives elsewhere. The full detail is retained in WPLOCATN.

- **CSPSBHH/INCSPSBHH** *(number of male R's nonbiological children (aged 18 or younger) in household who are the biological children of his current wife or cohabiting partner)* – This Section A recode was topcoded at "1 or more children under 19 in household" for public use. The full detail is retained in INCSPSBHH. See recode specifications in **Appendix 3c**.

- **NUMLIFE/ INNUMLIFE** *(female sexual partners in lifetime)*– The raw variable FE-1 NUMLIFE was topcoded at "50 or more female partners in lifetime" for public use. The full detail is retained in INNUMLIFE.

- **PARTS1YR/ INPARTS1YR** *(number of opposite-sex sexual partners in last 12 months)* – This male recode was topcoded at "7 or more female partners in last 12 months" for public use. The full detail is retained in INPARTS1YR. See recode specifications in **Appendix 3c**.

- **Summary variables of number of coresidential children in the household --** These computed variables from Section G (NUMCRU18, CRALL, CRALLU5, CRALL518, CRMALU5, CRMAL518, CRFEMU5, and CRFEM518, all computed in Flow Check G-0) summarize the type of coresidential children living in the respondent's household. They describe the number of children living in the household by age group and gender. Each of these variables was top coded for public use. The original, full-detail variables, all named with "IN" prefixes, are available through the RDC.

- **Child support paid per year for non-coresidential children** (CHSUPPYR, GC-3a NCAMOUNT, GC-3b NCUNIT, YEARSUPP) – The raw variables NCAMOUNT and NCUNIT

describe the amount of child support the respondent paid for his non-coresidential children aged 18 or younger. NCAMOUNT contains the dollar amount paid, and NCUNIT gives the frequency in which the respondent gave this support -- weekly, monthly, yearly. NCAMOUNT and NCUNIT were combined to define YEARSUPP, the dollar amount that men contributed *per year*, and were applicable for all men who have children aged 18 or younger who do not live with them. Then, a categorical variable **CHSUPPYR** was defined for public use based on YEARSUPP. The categories for CHSUPPYR, when applicable, are:

0=NONE/NOT REPORTED
1=$1 - $3,000
2=$3,001 - $5,000
3=$5,001 - $9,000
4=MORE THAN $9,000

The original variables NCAMOUNT, NCUNIT, and YEARSUPP are available through the RDC.

- **STARTMIL/CMBEGMIL --** The raw variables for the month and year the respondent began active duty in the Armed Forces for a period of 6 months or more (JC-2m BEGMIL_M and JC-2y BEGMIL_Y) and the corresponding century month variable computed in Flow Check J-10 (cmbegmil) are not included on the public use file due to disclosure risk. A variable indicating the year (in 5-year groupings) when the respondent began active duty (**startmil**) was created for public use. The original, full-detail variables (BEGMIL_M/_Y and cmbegmil) are available through the RDC.

- **ENDMIL/CMENDMIL** -- The raw variables for the month and year the respondent separated from active duty (JC-3m ENDMIL_M and JC-3y ENDMIL_Y) and the corresponding century month computed in Flow Check J-12 (cmendmil) are not included on the public use file due to disclosure risk. A variable indicating the year (in 5-year groupings) when the respondent separated from the Armed Services (**endmil**) was created for public use. The original, full-detail variables (ENDMIL_M/_Y and cmendmil) are available through the RDC.

FEMALE RESPONDENT FILE NOTES

LSTGRADE (computed variable defined in Flow Check A-21b):

LSTGRADE is "last elementary/junior high/high school grade attended". In the Year 1 female questionnaire instrument only, there were 85 respondents who skipped the necessary questions erroneously, so LSTGRADE=97 (not ascertained). Users may assign valid values by setting LSTGRADE equal to the value of AF-3 HIGRADE if it is non-missing, otherwise set it equal to AF-8 HISCHGRD if it is non-missing.

Marital Dissolution Data in Female CB series and Related Recodes (MARDISnn, MARENDnn):

In Cycle 6 (2002), there was a routing error in the CB series that resulted in higher levels of invalidly missing data on how marriages ended and what dates they ended. These routing errors were all fixed in the 2006-2010 NSFG, and the numbers of imputed values on MARDISnn and MARENDnn are very small.

Section C, CE series and subsequent:

7 cases responding "96" on WNFSTSEX_M (month of first sexual intercourse with a man), which means "never had sex," who also had ever been pregnant, married, or cohabited: These were not edited anywhere in the file so that the recode HADSEX is also "no" for these 5 cases. These cases were routed through the questionnaire as never having had sex, so variables that are relevant to pregnancies and sexual partners that occur after CE-1 EVERSEX will be system missing, or inapplicable. A special code was created for VRY1STSX, date of first sexual intercourse, for these cases: 9996 (this is also their value on the computed variable for date of first sex: cmfstsex). On the pregnancy file, the two of these cases with pregnancies will have "not ascertained" on variables that depend on information about first sex, that are missing for these cases. An example of this is computed variable "cmintstr."

Incorrect routing past questions on last partner characteristics (CI series):

For Years 1 and 2 of the questionnaire, for respondents whose last sexual partner was longer ago than 12 months (parts12=0), the following questions were skipped erroneously. If the last partner was also the first partner or a husband or cohabiting partner this information is available in earlier series. Otherwise it is missing. This is missing for approximately 360 cases. This was fixed starting with the Year 3 questionnaire.

CI-9 P1YRAGE, P1YRAGE2, P1YRAGE3
CI-10 P1YHSAGE, P1YHSAGE2, P1YHSAGE3
CI-11 P1YRFX, P1YRFX2, P1YRFX3
CI-12m P1YFSEX_M, P1YFSEX_M2, P1YFSEX_M3
CI-12y P1YFSEX_Y, P1YFSEX_Y2, P1YFSEX_Y3
cmfsexx, cmfsexx2, cmfsexx3

Open-ended questions for reasons for discontinuing specific contraceptives, EA series (added in Year 2, that is, starting in July 2007):

The following variables are a result of coding responses to open-ended questions (3 questions for each of 5 methods). One of the open-ended questions, a standard "other-specify" question, was in place throughout the 2006-2010 NSFG, and the other two were added during Year 2 of 2006-10. See **Appendix 5** (summary of questionnaire changes) for more details.

STOPPILL1-STOPPILL4
STOPCOND1-STOPCOND2
STOPDEPO1-STOPDEPO5
STOPPTCH1-STOPPTCH3
STOPLUNL1

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-21 REASLUNL, EA-22 REASPTCH. 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)."

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

EA_15 SP_OTHRMETH is the question capturing 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 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."

ED Series: Contraceptive Method History:

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 July 2006, the series collects method use information from January 2003 through July 2006. 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.

If interview year is **2006** (cmintvw=**1278** through **1284**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2003	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2004	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2005	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2006	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 **2007** (cmintvw=**1285** through **1296**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2004	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2005	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2006	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2007	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 **2008** (cmintvw=**1297** through **1308**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2005	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2006	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2007	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2008	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 **2009** (cmintvw=**1309** through **1320**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2007	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2008	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2009	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 **2010** (cmintvw=**1321** through **1326**), the variable names correspond to months/years as below:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2007	METHX1-4	METHX5-8	METHX9-12	METHX13-16	METHX17-20	METHX21-24	METHX25-28	METHX29-32	METHX33-36	METHX37-40	METHX41-44	METHX45-48
2008	METHX49-52	METHX53-56	METHX57-60	METHX61-64	METHX65-68	METHX69-72	METHX73-76	METHX77-80	METHX81-84	METHX85-88	METHX89-92	METHX93-96
2009	METHX97-100	METHX101-104	METHX105-108	METHX109-112	METHX113-116	METHX117-120	METHX121-124	METHX125-128	METHX129-132	METHX133-136	METHX137-140	METHX141-144
2010	METHX145-148	METHX149-152	METHX153-156	METHX157-160	METHX161-164	METHX165-168	METHX169-172	METHX173-176	METHX177-180	METHX181-184	METHX185-188	METHX189-192

Female Section F: Mapping of family planning variables

In female Section F we ask about a set of 14 different possible services women could have received in the past 12 months. For each service received we gather additional information 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 2006-2010 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	BC12PLCX	REGCAR12	BC12PAYX11-13
Checkup for Birth Control	FA-1c	MEDTST12	FPTITCHK	BC12PLCX 2	REGCAR13	BC12PAYX21-23
Counseling About Birth Control	FA-1d	BCCNS12	FPTITCBC	BC12PLCX 3	REGCAR14	BC12PAYX31-33
Sterilizing operation	FA-1e	STEROP12	FPTITSTE	BC12PLCX 14	REGCAR25	BC12PAYX141-142
Counseling re Getting Sterilized	FA-1f	STCNS12	FPTITCST	BC12PLCX 4	REGCAR15	BC12PAYX41-43
Emergency Contraception/Prescription	FA-1g	EMCON12	FPTITEC	BC12PLCX 5	REGCAR16	BC12PAYX51-52
Counseling re Emergency Contraception	FA-1h	ECCNS12	FPTITCEC	BC12PLCX 6	REGCAR17	BC12PAYX61-62
Pregnancy Test	FA-3a	PRGTST12	FPTITPRE	BC12PLCX 7	REGCAR18	BC12PAYX71-73
Abortion	FA-3b	ABORT12	FPTITABO	BC12PLCX 8	REGCAR19	BC12PAYX81-82
Pap Smear	FA-3c	PAP12	FPTITPAP	BC12PLCX 9	REGCAR20	BC12PAYX91-93
Pelvic Exam	FA-3d	PELVIC12	FPTITPEL	BC12PLCX 10	REGCAR21	BC12PAYX101-103
Prenatal Care	FA-3e	PRENAT12	FPTITPRN	BC12PLCX 11	REGCAR22	BC12PAYX111-113
Post-Pregnancy Care	FA-3f	PARTUM12	FPTITPPR	BC12PLCX 12	REGCAR23	BC12PAYX121-122
Counsel/test/treatment for STD	FA-3g	STDSVC12	FPTITSTD	BC12PLCX 13	REGCAR24	BC12PAYX131-133
All services in one visit:	FA-4	NUMBCVIS = 1 all in one visit	-	bc12plcx	regcar12	BC12PAYX11-13

Section J (ACASI): Erroneous routing past JD-2 AGEVAGR and JD-3 AGEVAGM, and effects for ages 15-17:

Flow Check J-7c was added to the questionnaire after Cycle 6 so that women who had ever been married, cohabited, or had ever been pregnant were skipped past the question asking whether she had ever had vaginal intercourse with a male (JD-1 VAGSEX). This skip was added to avoid unnecessary contradiction of their CAPI reports. An inadvertent consequence of this new flow check was that such respondents were also skipped past two questions asking about their age and their partner's age at this first sexual intercourse (JD-2 AGEVAGR, JD-3 AGEVAGM). For females aged 15-17 this is the only point in the NSFG where they are asked about age of their first male partner (in contrast to those 18 – 44, for whom it is also asked earlier in the CAPI portion in Section C). Minors were only asked about age of first partner in this series because it was the only acceptable way to obtain information on a male sexual partner who was potentially an adult. This improvement in questionnaire routing in Flow Check J-7c, then, inadvertently resulted in missing data for 127 females aged 15-17 on age of first male sexual partner. This is 32% of sexually experienced females age 15-17, thus analysis of this subgroup is adversely affected.

The recode that represents this information is FSEXPAGE (Age of R's 1st sexual partner at time of R's 1st sex). As is standard for recodes, missing values for all cases, including these minors, have been imputed.

Since the age of the first male partner is reported by the total age range of females up to 44, analysts may wish to use the reports of females age 18 and over, whose first sex was at age 17 and under, or any age of interest within that range, for reliable information on age of first male partner for minors.

For women aged 18-44, the intent of asking for age at first sex and age of partner at first sex again in ACASI was to allow comparison between CAPI and ACASI reports on these measures. This capability, and the age of first partner for minors, will be restored with years of NSFG data collection beyond 2010.

Recode LSEXRAGE: Respondents with values of 45 years old:

On the female recode LSEXRAGE (respondent's age at last sexual intercourse), 9 respondents have a value of 45 years. Four 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 before their actual birthday in that month. However, the formula for computing LSEXRAGE yielded 45 years because it is calculated as follows: century-month of last sexual intercourse minus century-month of respondent's birth, divided by 12.

LSEXPAGE: Reverted from recode to "computed in post-processing" due to imputation challenges:

This variable was originally intended to be a recode. However, a routing error caused a

larger-than-expected amount of missing data on a key input variable (CI-10 P1YHSAGEX) (see "Incorrect routing past questions on last partner characteristics (CI series)" in this User's Guide). It was judged too risky to impute this amount of data with the potential to introduce a large number of unrealistic age combinations of respondent and partner at last sex. Therefore the missing data were not imputed, and assigned "997" for "not ascertained." The following are the specifications for the construction of this variable:

LSEXPAGE is blank (inapplicable) if R has never had sexual intercourse (recode HADSEX=2).

If R's most recent partner was R's first partner ever (LIFEPRT=1 or MATCHFP=1 or p1yrelp=13 or (p1yrelp=7 and CG-2 SAMEMAN=1 and CA-1 TIMESMAR=1) or (p1yrelp=8 and CG-2 SAMEMAN=1 and TIMESCOH=1)):

If R had sex only once with first partner (cmlsxfp=9996) then assign partner's age at first sex with R:
LSEXPAGE = recode **FSEXPAGE**

Else, determine # of months that elapsed between first and last sex with first partner:
calculate elapsed1:
If cmlsxfp not sysmis, 9998, or 9999 and cmfstsex not sysmis, 9998, or 9999 then:
elapsed1 = cmlsxfp minus cmfstsex

Then, estimate partner's age at last sex based on elapsed1 value and partner's age at first sex with R, recode **FSEXPAGE**, and flag with leading 9:

LSEXPAGE = **FSEXPAGE** + INT(elapsed1/12)) + 900

Else if R's most recent partner was <u>not</u> R's first partner ever (LIFEPRT>1 and p1yrelp ne 13):

[Note: for those age 18 or over, and those under 18 but partner not current, all we have is partner's age at <u>first</u> sex with R. Use Section C variable for partner's age at first sex, and if had sex more than once with him, add the number of months elapsed between first & last sex to this age. For those under 18 whose partner is current, all we have is partner's age at <u>last</u> sex with R. Use Section J variable for partner's age at last sex with R (doesn't matter whether had sex with him more than once)]

If R is 18 or older (age_r ge 18) or
R is under age 18 (age_r < 18) and partner is not current (CI-7 P1YCURRPX ne 1 or parts12=0), then use age of most recent partner at first sex with R, from Section C:

If R had sex only once with most recent partner (cmfsexx=9996) then assign (valid) partner's age at first sex with R:

if 5 le CI-10 P1YHSAGEX le 95 then:

LSEXPAGE = CI-10 P1YHSAGEX

Else, determine # of months that elapsed between first and last sex with most recent partner:
calculate elapsed2:
If cmfsexx not sysmis, 9998, or 9999 and cmlsexx not sysmis, 9998, or 9999 then:
elapsed2 = cmlsexx minus cmfsexx

Then, estimate partner's age at last sex based on elapsed2 value and (valid) partner's age at first sex with R, and flag with leading 9:

if 5 le CI-10 P1YHSAGEX le 95 then:
LSEXPAGE = CI-10 P1YHSAGEX + INT(elapsed2/12)) + 900

Else if R is under age 18 (age_r < 18) and partner is current (CI-7 P1YCURRPX=1), then use (valid) age of most recent partner at last sex, from ACASI, section J:

If JF-2a CURRPAGE <= 95 then LSEXPAGE=KG-3a CURRPAGE

Else if JF-2a CURRPAGE = DK/RF then:
[If partner was 1-2 years older, add 2 years to R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=1 and JF-2c HOWMUCH=1) then
LSEXPAGE=(VRY1STAG + 2) + 900
[If partner was 3-5 years older, add 4 years to R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=1 and JF-2c HOWMUCH=2) then
LSEXPAGE=(VRY1STAG + 4) + 900
[If partner was 6-10 years older, add 8 years to R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=1 and JF-2c HOWMUCH=3) then
LSEXPAGE=(VRY1STAG + 8) + 900
[If partner was more than 10 years older, add 10 years to R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=1 and JF-2c HOWMUCH=4) then
LSEXPAGE=(VRY1STAG + 10) + 900

[If partner was 1-2 years younger, subtract 2 years from R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=2 and JF-2c HOWMUCH=1) then
LSEXPAGE=(VRY1STAG - 2) + 900
[If partner was 3-5 years younger, subtract 4 years from R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=2 and JF-2c HOWMUCH=2) then
LSEXPAGE=(VRY1STAG - 4) + 900
[If partner was 6-10 years younger, subtract 8 years from R's age at 1st sex and flag with leading 9]
if (JF-2b RELAGE=2 and JF-2c HOWMUCH=3) then

LSEXPAGE=(VRY1STAG - 8) + 900

[If partner was more than 10 years younger, subtract 10 years from R's age at 1st sex and flag with leading 9]

if (JF-2b RELAGE=2 and JF-2c HOWMUCH=4) then LSEXPAGE=(VRY1STAG - 10) + 900

[If partner was the same age, then use R's age at last sex and flag with leading 9]

if JF-2b RELAGE=3 then LSEXPAGE=VRY1STAG+ 900

Notes:

- *Under some conditions, computed based on imputed values of FSEXPAGE.*
- *Where the "don't know followup" questions RELAGE and HOWMUCH are used: in some cases they have valid data but were not used because the combinations of values didn't meet the criteria above.*

Code categories:

Blank = Inapplicable
00-95 = Partner's age at last sexual intercourse, reported
900-995=Partner's age at last sexual intercourse, estimated
997 = Not ascertained

FEMALE PREGNANCY (INTERVAL) FILE NOTES

HPAGEEND (father's age when a nonlivebirth pregnancy ended) omitted from file:

Due to a programming error, BC-4c HPAGEEND was not asked for all applicable pregnancies as it should have been for all pregnancies that did not end in live birth. As a result of the higher levels of illegitimately missing data, the variable was not included on the pregnancy file. BD-6 HPAGELB, which asks the age of the father when each live birth occurred, was asked correctly, and is available on the pregnancy file.

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

In the 2006-2010 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 2006-2010, a somewhat different approach was taken to sorting and correcting the pregnancy-specific information that women reported, and it was adopted 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)	wksgestv	PRGLNGTH

Century month for date of beginning of pregnancy interval - cmintstr:

Two cases have a value of "97," for "not ascertained" on the computed variable cmintstr, defined in Flow Check E-68. The beginning of the first pregnancy interval is the date of first sexual intercourse. These cases responded "96" on WNFSTSEX_M (month of first sexual intercourse with a man), which means "never had sex," but they had ever been pregnant, and provided information on their pregnancies in the interview (also see FEMALE RESPONDENT FILE NOTES item on "Section C, CE series and subsequent").

MALE FILE NOTES

Biological children in chronological order:

In the male questionnaire, respondents were asked about their biological children in the context of questions about the mothers of these children. Based on consultation 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 2006-2010 NSFG, there were never more than 10 children reported across these woman-specific loops in Sections C-E or in Section F. See exact numbers below:

- **Section C** – space for 10 children, but no respondent reported more than 8
- **Section D** – space for 10 children for each partner from this section, but:
    - No more than 7 reported with "P1" most recent partner
    - No more than 5 reported with "P2" 2nd-most-recent partner in past year
    - No more than 3 reported with "P3" 3rd-most-recent partner in past year
- **Section E** – space for 10 children with each of up to 10 former wives and with 1st cohabiting partner, but only 3 former wives were reported in the 2006-2010 NSFG and:
    - No more than 6 reported with "W1" 1st former wife
    - No more than 3 reported with "W2" 2nd former wife
    - No more than 3 reported with "W3" 3rd former wife
    - No more than 5 reported with "1st Cohab Partner"
- **Section F** – space for 10, but no respondent reported more than 9

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 15 characteristics listed below are arranged as arrays of 10 variables each because just as in Cycle 6 (2002), no male respondent reported fathering more than 10 biological children in the 2006-2010 NSFG.

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

1. BIODOB1-10: Century month of child's birth
2. BIOSEX1-10: Sex of child

3. BIOAGE1-10: Age of child
4. BIOHH1-10: Whether child lives in same household with respondent
5. BIOMOM1-10: R's relationship to biological mother of this child *(this array is new for the 2006-2008 NSFG, and is defined based on what section of the questionnaire the child was reported in)*
6. BIOMAR1-10: Whether respondent was married to child's mother at time of child's birth
7. BIOCOHB-10: Whether respondent was living with child's mother at time of child's birth (includes cohabiting or married)
8. BIOLRNPG1-10: When respondent learned of the pregnancy (before or after child was born)
9. BIOLGPAT1-10: Whether legal paternity was established (if unmarried at child's birth)
10. BIOHSPAT1-10: Whether paternity was established at the hospital
11. BIOLVEVR1-10: Whether respondent ever lived with child (if not living with child now)
12. BIOHWFAR1-10: How far away child lives (in miles) from respondent
13. BIOWANT1-10: Wantedness of the pregnancy by respondent
14. BIOHSOON1-10: Timing of the pregnancy
15. BIOHPYPG1-10: Respondent's happiness about the pregnancy

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 documentation 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
BIODOB1-10	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]
BIOMULT1-10	Whether part of a multiple birth	CG-7 MULTBIRT	DH-7 MULTBIRT	ED-7 MULTBIRT	FA-10 MULTBIRT
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 CP	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 FW)	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
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 3 mentions for each child)	CG-11 CWPCHLIV	DH-11 PXCXLIV	ED-11 FWPCHLIV	FA-14 OBCLIV
BIOLGPAT[x]	Has legal paternity been established (if unmarried at child's birth)	CG-13 CWPCHLEG	DH-13 PXCXLAW	ED-13 FWPCHLEG	FA-16 OBCLAW
BIOHSPAT[x]	Was paternity established at hospital	CG-14 CWPCHHOP	DH-14 PXCXHOP	ED-14 FWPCHHOP	FA-17 OBCHOP
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

Nonbiological children cared for by men (Sections CI, CJ, DJ, DK, EE, and EF):

In Year 1 of the 2006-2010 NSFG, which began interviewing in June 2006, men were asked for the number of nonbiological children he had cared for 1) within each of his marriages or cohabitations and 2) outside of any relationship with a female partner. However, he was only asked further details about the child (sex, age, and current living situation) if he had adopted the child. Given that men do not formally adopt the majority of nonbiological children they care for with their female partners, this questionnaire routing resulted in a significant loss of information about men's parenting of nonbiological children.

In order to capture information about all the nonbiological children for whom the man cared, beginning in Year 2 (July 2007) information began to be collected in Section C for all children of his current wife or partner who had lived with the respondent and all children not related to him or his current wife or partner who had lived with them. In addition, men were asked whether they adopted each of these children.

In Year 3 (July 2008), information about children of former wives or cohabiting partners began to be collected in Sections D and E. See summary table for further details.

Questionnaire section	Collected in Year 1	Collected in Year 2	Collected in Year 3
CI. Current wife's or partner's children from previous relationships	Adopted children	Nonbiological children who lived with the man	Nonbiological children who lived with the man
CJ. Other nonbiological children with current wife or partner	Adopted children	Nonbiological children who lived with the man	Nonbiological children who lived with the man
DJ. Recent former wives' or partners' children	Adopted children	Adopted children	Nonbiological children who lived with the man
DK. Other nonbiological children with recent former wife or partner	Adopted children	Adopted children	Adopted children
EE. Other wives' or partners' children	Adopted children	Adopted children	Nonbiological children who lived with the man
EF. Other nonbiological children with other wives or partners	Adopted children	Adopted children	Adopted children

When information was captured for all nonbiological children, the specifications to compute the variables that count the number of children a man adopted, e.g., C_OKAKIDS, were modified accordingly.

Contraceptive method use variables in the male questionnaire and data file:

The male questionnaire contains questions asking for specific contraceptive methods used at first and last sexual intercourse with current wife or cohabiting partner (Section C) and with each of up to 3 partners in the past 12 months (Section D). Guidance is given here because these question series underwent multiple changes from Cycle 6 to Year 1 of 2006-2010 and from Year 1 to Year 2 of 2006-2010:

- These sets of questions were involved in an experiment that was discontinued after Year 1;
- The variables are based on "code-all-that-apply" questions, so they have several iterations corresponding to each question;
- New response categories were added for additional contraceptive methods from Cycle 6 to 2006-2010, and from Year 1 to Year 2
- In the Section D series, they are repeated for up to 3 partners (since they are within the series on method use with nonmarital, noncohabiting partners in the past 12 months or last partner ever)

Questions on contraceptive method use in Sections C and D of the male questionnaire were involved in an experiment to test the effects of question wording, for Year 1 of the 2006-2010 release. This is a continuation of an experiment begun with Cycle 6, and details are available in the User's Guide for Cycle 6, pages 26 and 27. This experiment was discontinued beginning in Year 2 of interviewing for the 2006-2010 NSFG because enough data had been accumulated. Beginning in Year 2, all references to the experiment and the two sets of experimental questions are no longer needed and are absent from the CRQ and CAPI-lites. In the codebook, all references to the experiment and the experimental questions are present since the final variables reflect Year 1's data and subsequent. For all these variables, their "universe statements" contain the information necessary to understand the universes in both Year 1, during the experiment, and Years 2 and beyond, after its discontinuation.

See **Appendix 5** for a table of these questions and their descriptions in 2002 (Cycle 6), and Years 1, 2 and 3 of 2006-2010. These questions and variables did not change after Year 2.

The following recodes on the male public use file combine these sets of questions, and generate measures of contraceptive method use: LSEXUSE1-LSEXUSE4; METH12M1-METH12M4; METH3M1-METH3M4; SEX1MTHD1-SEX1MTHD4. See **Appendix 3c** for further information on how these recodes were defined. These recodes are consistent with Cycle 6 recodes with the same names.

Recode LSEXRAGE: Respondents with values of 45 years old:

On the male recode LSEXRAGE (respondent's age at last sexual intercourse), 10 respondents have a value of 45 years. Three 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 7 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 yielded 45 years because it is calculated as follows: century-month of last sexual intercourse minus century-month of respondent's birth, divided by 12.