Frequently Asked Questions (FAQ) about the
National Survey of Family Growth (NSFG)

List of Questions

1)   *How were data collected for the 2022-2023 NSFG?*

NSFG data for 2022-2023 were collected using a multimode design involving both in-person interviews and web survey completion. Data collection was conducted from January 2022 to December 2023 with a nationally representative sample of male and female respondents 15-49 years of age. Information on the 2022-2023 NSFG's survey design, sample design, new features, and weighting and variance estimation, can be found in the User's Guide sections "**Survey and Sample Design for 2022-2023**" and "**Sample Weights and Variance Estimation**." Further details on the survey design and operations for the 2022-2023 NSFG will be posted on the NSFG webpage in the following 4 documents (expected by Summer 2025):

- Summary of Design and Data Collection Methods
- Sample Design Documentation
- Sample Error Estimation Design
- Weighting Design Documentation

2) *How were people selected to participate in the 2022-2023 NSFG?*

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

3) *Why are there 3 different data files for 2022-2023 NSFG?*

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

- The <u>female respondent</u> file contains 5,586 records - one record per female ages 15-49 who completed the survey in 2022-2023.

- The <u>female pregnancy</u> file contains 8,247 records - one record per pregnancy reported by female respondents in 2022-2023. The pregnancy file includes each respondent's complete pregnancy history, regardless of the year in which the pregnancy occurred. If a female respondent has never been pregnant, she has no pregnancy file records; if she reports 5 pregnancies, she has 5 pregnancy records in the file. To reduce the need for merging, the female respondent file includes selected variables from the pregnancy file, and the pregnancy file includes selected variables from the respondent file.

- The <u>male respondent</u> file contains 4,371 records - one record per male ages 15-49 who completed the survey in 2022-2023. This file includes all information on births and other pregnancies fathered by men in the sample, but this information is presented with the man as the unit of observation; there is no "pregnancy-based" file for male NSFG respondents.

See section on "**Organization of the 2022-2023 NSFG Public-Use Data Files**" in the User's Guide; also see (posted on NSFG webpage) the **Public-Use File Indexes** for each file for full lists of all the variables contained in these three public-use data files and the **Lists of Restricted-use Analytic Variables** available through the NCHS Research Data Center (RDC).

4) *How do I download the 2022-2023 NSFG public-use data files?*

The female, pregnancy, and male public-use data files in CSV and SAS7BDAT formats are available for download on the website.

To download a data file:
Right-click on the file you want to download, then click on "Save Target As..." (Internet Explorer) or "Save Link As..." (Firefox)
(On macOS, click on the link with the option key held down.)

5) *How do I read the 2022-2023 NSFG data files into my statistical software packages?*

Data are provided in CSV and SAS7BDAT formats. These formats can be imported or read into various software packages such as SAS, R, Stata, and SPSS. The CSV files are accompanied by a ReadMe text file that details how to import the file into SAS and R. Users should implement these instructions when importing the file to avoid issues with reading the full data file.

6) *Where can I find the codebooks and questionnaires for the 2022-2023 NSFG?*

The 2022-2023 NSFG **codebooks** are accessible on the NSFG webpage, with 1 PDF for each of the 3 public-use data files. Please see the User's Guide section called **"Description of Codebooks"** for further information on the elements of each codebook entry: variable name, variable type, variable label, universe statements ("applicable specifications"), response categories and unweighted frequencies, and where relevant, special user notes.

All data collection for the 2022-2023 NSFG was conducted using computer-assisted personal interviewing (CAPI). The 2022-2023 NSFG **questionnaires** are available on the NSFG webpage in two formats – CAPI-lite, an abridged version that shows essential question wording and routing for the male and female surveys, and the CAPI Reference Questionnaire (CRQ) version that shows full specifications for the male and female surveys.

7) *Do I need to use the sampling weights? What sampling weights should I use for my analyses?*

Yes, it is essential to use the sampling weights when analyzing the NSFG data. The NSFG uses a multi-stage, probability-based complex sample design, not simple random sampling, to yield estimates representative of the U.S. household population aged 15-49. To control the costs of data collection and to obtain adequate sample sizes, the NSFG sampled some population groups at higher rates than others. As a result of this oversampling and other factors for which the sampling weights are adjusted, such as survey nonresponse, each respondent in the NSFG sample represents a different number of people in the U.S. household population. As in previous NSFG file releases, the sampling weight can vary significantly across respondents, so using the sampling weights is critical for producing accurate statistics.

The two-year sample weight (WGT2022_2023) is the final, fully adjusted weight that should be used in all analyses using the 2022-2023 NSFG sample of 4,371 male and 5,586 female

respondents who completed the survey over the 24-month period from January 2022-December 2023. In addition to using sampling weights, researchers must use the design variables for the sampling stratum (VEST) and cluster (VECL) to obtain correct standard errors for their estimates.

8)   *Why do you recommend using the recodes?*

Recodes in the NSFG are frequently used, "constructed" variables that NCHS has checked for consistency and **imputed** missing values in most cases. Many of the variables used in NCHS reports are recodes, and this makes it easier for researchers to replicate NCHS results. While some recodes are straightforward, others represent complex measures and can save work in your analyses. Imputed recodes are not available for all measures, but to the extent possible, researchers are urged to use the available recodes because they have been checked extensively and their missing values have been handled in a uniform manner. See the section on **"Recodes and Imputation"** in the User's Guide for more information. You will find a list of some commonly used recodes in that section. Also, in the public-use file codebooks and in the public-use file indexes, recodes are indicated in the "variable type." Recode specifications describing how each recode was defined and imputed (if imputed) are provided on the NSFG webpage as well.

9)   *Where can I find more information on how missing data are handled?*

For most variables missing data are discussed in several sections of the User's Guide main text, but particularly the sections on **"Coding for 'Don't Know,' 'Refused,' and 'Not Ascertained' Values"** and **"Century Month Coding for Dates."** Missing data are imputed for most recode variables (see Question #8).

10)   *What are the criteria used to assess the reliability of NSFG estimates in Key Statistics and NCHS reports?*

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

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

The best way to see the skip patterns or routing for questions in the survey is by looking at the questionnaires posted on the webpage. The questionnaires are provided in two formats (as described in Question #6): the full specifications laid out in the CRQs show detailed skip patterns, question wording variants, and consistency checks, while the more streamlined CAPI-

lites show an abridged version of skip patterns. See the User's Guide "**Description of Questionnaires**" for further details. In addition to the questionnaires, each variable's codebook entry includes a "universe statement" (also known as applicable specification) that indicates the set of cases for which the variable is asked or defined. These may be as simple as "Applicable for all respondents" or "Applicable if respondent is 15-24 years old," or they may be more complex. For further information, see **"Universe Statements ("Applicable Specifications")"** within the User's Guide section called **"Description of Codebooks."**

12) ***Can I combine data from different NSFG data files? Can estimates from 2022-2023 NSFG be compared with estimates from prior releases?***

The **2011-2019 Combined Files: Selected Data and Documentation** webpage provides technical guidance and appropriate weight variables for combining data across the four 2-year file releases from 2011-2019. Those weights were designed to represent the national population over the various 4-, 6-, and 8-year fieldwork periods across this continuous span of data collection and permit estimation of population totals as well as percentages and other statistics. However, due to the break in NSFG data collection from October 2019 through December 2021 (which included the early years of the COVID-19 pandemic), as well as the change to a multimode survey design for the 2022-2023 NSFG, weights were not produced for combining 2022-2023 data with prior file releases. Given the discontinuity of data collection and changes in survey design, *NCHS does not advise that users combine data* for 2022-2023 with prior NSFG data releases for purposes of making populations estimates for a wider span of years. In addition, we urge caution when *comparing* estimates from 2022-2023 to prior NSFG estimates (for example, for 2017-2019). Any differences seen (or lack thereof) across time may be confounded by change to the multi-mode survey administration, lower response rates, and other survey methodology factors all of which may be interrelated. See User's Guide section on "**Survey and Sample Design for 2022-2023**" for further details.

13) ***How has the NSFG questionnaire changed since 2017-2019?***

A **summary of questionnaire changes since 2017-2019** is posted on the NSFG webpage. The **questionnaires** posted on the NSFG webpage also document any changes that were made during the 2022-2023 data collection period.

14) ***Why are certain variables not included (or not included with the level of detail shown in the questionnaires) on the public-use file?***

NCHS has a legal and ethical mandate to prevent disclosure of the identities of NSFG respondents. To honor that mandate and make as much of the survey data available publicly as possible for researchers, NSFG staff take a significant set of disclosure-risk-reduction actions before releasing these public-use data files. The files also undergo intensive review by the NCHS Disclosure Review Board before they are approved for release. These risk reduction actions include the suppression or modification of several variables that could

permit indirect identification of survey respondents, particularly if used in conjunction with other data sources.

Please consult the User's Guide section on "**Protections to Minimize Risk of Disclosure for Individual-Level Data**" for further details. As described there, the codebook entries for those variables that have been modified, and in some cases created, for public use will include a special note. In addition, these variables have been noted with an asterisk in the Public-Use File Indexes provided on the webpage**.** Some of the original, full-detail variables that may be of analytic value for researchers are available as restricted-use data through the NCHS Research Data Center (RDC). These **listings of restricted-use analytic variables** available through the RDC are posted on the webpage.

If your research requires the use of restricted-use variables, visit the RDC website for information on the application process. You may also contact the NSFG team at nsfg@cdc.gov if you have questions about using NSFG data in the RDC. There are fees associated with using the RDC.

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

No. Although each year of NSFG data collection is designed to be nationally representative of the U.S. household population, sample sizes for a single year are too small to provide estimates with adequate levels of precision. Two years is generally the shortest time frame yielding enough numbers of cases and statistical stability. Based on this, weights are not provided for single years of NSFG data collection. If you have questions about analyzing changes over time by including covariate(s) representing specific years, email nsfg@cdc.gov.

16)   *Can the NSFG be used to produce state-level estimates?*

No. The NSFG sample is not designed to produce state-level (or lower geographic-level) estimates. The survey was designed to create national estimates. State-level characteristics, region of residence (4 categories of Northeast, South, Midwest, and West), and other contextual data can be used with the data files in NCHS's Research Data Center. Please see the Research Data Center webpage for information on accessing geographic and contextual data. The recode METRO is available on the public-use files. This variable categorizes the respondent's place of residence at the time of interview in 3 groups: principal city of a metropolitan statistical area (MSA), other MSA, or not an MSA. These 3 categories can be collapsed to create an indicator of urban-rural residence.

17)   *Does the display of data values in the codebook reflect the actual values of the variables in the data?*

The values of continuous variables such as numbers of partners, year of an event, or age at an event have sometimes been combined into groups of values to shorten the length of the

codebook entry. However, the separate values are still retained and accessible in the public-use data files. For example, if you see something like "15-19 = 15-19 years" in the codebook entry, this indicates that separate values of 15 through 19 for this variable are still contained in the data file, but they have been combined for codebook display purposes only. In contrast, if a variable's codebook entry shows "1 = 15-19 years," this means it is a categorical variable where the value 1 represents this group of ages.

18) *I have reviewed the documentation on the webpage and still have a question. Where can I get help?*

If you have reviewed the **User's Guide and other documentation** on the webpage and *cannot find the answer to your question, contact the NSFG team via email at* nsfg@cdc.gov.