Introduction to National Survey of Family Growth (NSFG)

Hsueh-Sheng Wu
CFDR Workshop Series
July 18, 2016



Outline

- Aims of NSFG
- NSFG study design
- NSFG data files
- Examples of faculty's work using NSFG
- Navigate NSFG website
 - 2011-2013 NSFG webpage
 - Questionnaire
 - Interactive codebook
 - Download data
- Weighting in NSFG
- Sample Stata codes for:
 - Reading NSFG data file into statistical software
 - Linking a female respondent file with a pregnancy file
 - Pooling men and women data from the same cycle
 - Using weight variables in analyses
- Conclusions



Aims of NSFG

- Address Sec 306 of the Public Health Service Act: "NCHS shall collect statistics on... family formation, growth, and dissolution."
 - Provide reliable national data on marriage, divorce, sexual activity, contraception, pregnancy, sterilization, and infertility.
 - Describe attitudes of sex, marriage, cohabitation, and parenthood.
 - Monitor health risks, sexually transmitted diseases (including HIV), and the health of infants, women, and men.
 - Policies can be made to meet the needs of families
 - Service agencies can design health and service programs



NSFG Study Design

| Name | Year Target population: Non-institutionalized people in the U.S. | | Survey Design (repeated cross-sectional | N | | |
|--------------------------------------|--|--------------|---|---|--------------------------------------|--------------------------------|
| | | Gender | Age Range | marital or parental status | | |
| Cycle 1 | 1973 | women | 15-44 | ever married or have offspring in the household | Interviews completed in a year | 9,797 |
| Cycle 2 | 1976 | women | 15-44 | ever married or have offspring in the household | Interviews completed in a year | 8,611 |
| Cycle 3 | 1982 | women | 15-44 | | Interviews completed in a year | 7,969 |
| Cycle 4 | 1988 | women | 15-44 | | Interviews completed in a year | 8,450 |
| Cycle 5 | 1995 | women | 15-44 | | Interviews completed in a year | 10,847 |
| Cycle 6 | 2002 | women men | 15-44 | | Interviews completed in a year | 12,571 W=7,643 M=4,928 |
| 2006-2010 NSFG | 2006-2010 | women men | 15-44 | | Interviews spread over several years | 22,682 W=12,279 M=10,403 |
| 2011-2013 NSFG | 2011-2013 | women men | 15-44 | | Interviews spread over several years | 10,416 W=5,601 M=5,815 |
| 2015-2019 NSFG Family a Demogr | 2015-2019 nd | women men | 15-49 | | Interviews spread over several years | n/a |

NSFG Study Design (Continued)

| Name | Year | Sampling Method | Over-Sampled groups | Data Collection Methods |
|----------------|-----------|---|----------------------------|--|
| Cycle 1 | 1973 | multistage probability sampling | Black women | Personal interview |
| Cycle 2 | 1976 | multistage probability sampling | Black women | face-to-face interview |
| Cycle 3 | 1982 | multistage area probability sampling | Black women & Teens | face-to-face interview |
| Cycle 4 | 1988 | Using Households used in National Health Interview Survey | Black women | face-to-face interview |
| Cycle 5 | 1995 | Using Households used in National Health Interview Survey | Black and hispanic women | Computer-assisted personal interview (CAPI) |
| Cycle 6 | 2002 | multistage area probability sampling | Blacks, Hispanics, & Teens | Interviewer administered CAP self-administered ACAPI |
| 2006-2010 NSFG | 2006-2010 | multistage area probability sampling | Blacks, Hispanics, & Teens | Interviewer administered CAP self-administered ACAPI |
| 2011-2013 NSFG | 2011-2013 | stratified multi-stage area probability sampling | Blacks, Hispanics, & Teens | Interviewer administered CAP self-administered ACAPI |
| 2015-2019 NSFG | 2015-2019 | n/a | n/a | n/a |



NSFG Data Files

Each NSFG survey has at most five types of data files:

- Public-use data were collected through personal interviews or Computer Assisted Personal Interviewing (CAPI)
 - Respondent data were information about female or male respondents
 - Interval data were information about pregnancies of female respondents
 - → This is the primary data that users work with
- ACASI data were collected in the Audio Computer-Assisted Self Interviewing (ACASI) portion of the interview.
 - ACASI data collect information on sensitive topics, such as alcohol and drug use, sexual risk behaviors, non-voluntary sexual intercourse, and experiences of oral sex, anal sex, and same-sex sexual partners, among other things
 - ACASI data for 2011-2013 are already included in the public-use data file.
 However, to apply for ACASI data from Cycles 5, 6, and 2006-2010,
 researchers need to send their request to NSFG staff (nsfg@cdc.gov).



NSFG Data Files (Continued)

- Contextual data: Geographic variables were created for the respondent's address.
 - These variables are available for Cycles 5, 6, 2006-2010 NSFG, and 2011-2013 NSFG. NSFG surveys at different years may have different levels of contextual variables. For example, Cycles 5, 6 and 2006-2010 NSFG have context variables measured at the state, county, tract, block group, and block levels, but 2011-2013 NSFG have variables at the state and county levels.
 - To obtain the files, researchers need to contact the Research Data Center (RDC) at NCHS
- Region data: One region variable indicates whether the respondent lives in Northeast, Midwest, South, or West.
 - Region variable is available for 2011-2013 NSFG.
 - To obtain the region data, researchers need to send request to NSFG staff (nsfg@cdc.gov).



NSFG Data Files (Continued)

- Interviewer observation: When respondents complete the ACASI portion of survey, interviewers filled out a paper questionnaire about the interview process
 - These data are used mainly for survey planning and management, for evaluating data quality, and for assessing the effectiveness of interviewer training.
 - Interviewer observation data file is also available through RDC



Availability of NSFG Data Files

| Table 2. NSFG dat | a files | | | | | |
|-------------------|-----------|--|------------|-----------------|-------------|-----------------------------------|
| Name | Year | Public-Use Data Fiile | ACASI data | Contextual data | Region Data | Interviewer observational data |
| Cycle 1 | 1973 | 1973NSFGData.dat | | | | |
| Cycle 2 | 1976 | 1976NSFGData.dat | | | | |
| Cycle 3 | 1982 | 1982NSFGData.dat | | | | |
| Cycle 4 | 1988 | 1988FemRespData.dat 1988PregData.dat | | | | |
| Cycle 5 | 1995 | 1995FemRespData.dat 1995PregData.dat | × | х | | X |
| Cycle 6 | 2002 | 2002FemResp.dat 2002FemPreg.dat 2002Male.dat 2002curr_ins.dat 2002HHvars.dat | х | х | | х |
| 2006-2010 NSFG | 2006-2010 | 2006_2010_FemResp.dat 2006_2010_FemPreg.dat 2006_2010_Male.dat | х | Х | | x |
| 2011-2013 NSFG | 2011-2013 | 2011_2013_FemRespData.dat 2011_2013_FemPregData.dat 2011_2013_MaleData.dat | Х | Х | х | х |
| 2015-2019 NSFG | 2015-2019 | n/a | n/a | n/a | n/a | n/a |



Topics in 2011-2013 NSFG

| | Male Respondent Data | Female Respondent data |
|------------|--|---|
| | Respondent ID and Selected Screener Items | Respondent ID and Pregnancy Order |
| Section A. | Demographic Characteristics; Household Roster; Childhood Background; Marital/Cohabiting Status | |
| Section B | Sex Communication, Ever Sex, Number of Sexual Partners, Enumeration and Relationship With Up To 3 Recent (Or Last) Sexual Partner(s) | Raw variables |
| | | Pregnancy-based recodes |
| Section C | Current Wife or Cohabiting Partner | |
| Section D | Recent (Or Last) Sexual Partner(s) and First Sexual Partner | |
| Section E | Former Wives and First Cohabiting Partner | Raw variables |
| | | Pregnancy-based recodes |
| Section F | Other Biological Children, Other Adopted Children, Other Pregnancies | |
| Section G | Fathering | |
| Section H | Desires and Intentions for Future Children | |
| Section I | Health Conditions and Health Services | |
| Section J | Residence and place of birth; Religion; Military service; Past and current work (R and current wife/partner): Attitudes | |
| Section K | Audio CASI | |
| | Recode Variables and Imputation Flags | Selected respondent file variables |
| | Weights and Other Variables | Pregnancy recode imputation flags |
| | | Respondent recode imputation flags included or pregnancy file |
| amily | and | Weights and related variables |
|)emoi | graphic Research | Date of interview, key reference dates, and fieldwork variables |

Topics in 2011-2013 NSFG (Continued)

| | Pregnancy Data file |
|------------|---|
| | Respondent ID and Selected Screener Variables |
| Section A. | Calendar Instructions; Demographic Characteristics; Household Roster; Childhood Background |
| Section B | Pregnancy & Birth History; Adoption & Nonbiological Children |
| Section C | Marital and Relationship History |
| Section D | Sterilizing Operations and Impaired Fecundity |
| Section E | Contraceptive History and Pregnancy Wantedness |
| Section F | Family Planning and Medical Services |
| Section G | Birth Desires and Intentions |
| Section H | Infertility Services and Reproductive Health |
| Section I | Insurance; Residence and place of birth; Religion; Past and current work (R and current H/P); Attitudes |
| Section J | SECTION J: Audio CASI |
| | Recode Variables and Imputation Flags |
| | Weights and Other Variables |

11

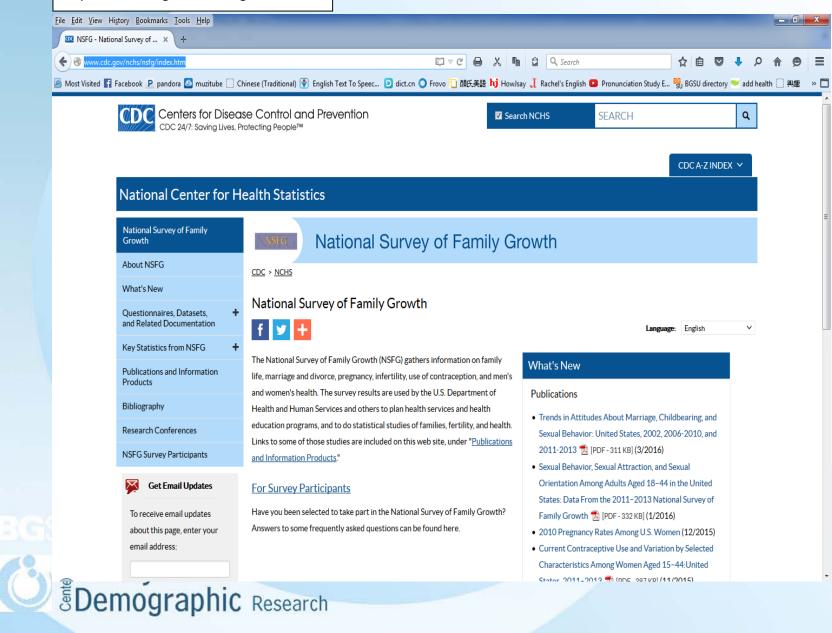
Examples of Faculty's Work using NSFG

- Guzzo K. B., & Hayford, S. (2011). Fertility following an unintended first birth. *Demography*, 48, 493-516.
- Joyner K, Peters, H.E., Hynes, K, Sikora, A, Taber, J.R., & Rendall, M.S. (2012) The quality of male fertility data in major U.S. surveys. *Demograph*y,49,101-122.
- Manning, W. D., & Cohen, J. A. (2012) Premarital cohabitation and marital dissolution: An examination of recent marriages, *Journal of Marriage and Family*, 74, 377–387.
- Manning W.D., & Brown, S. L. (2014). Two decades of stability and change in age at first union formation. *Journal of Marriage and Family*, 76, 247-260.
- Stykes, J. B., Manning, W. D., & Brown, S. L. (2013) Nonresident fathers and formal child support: Evidence from the CPS, NSFG, and SIPP. *Demographic Research*, 29,1299-1330

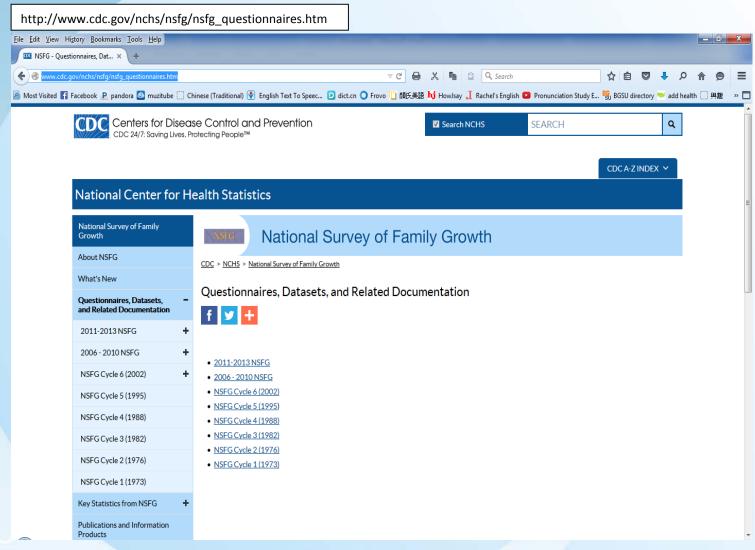


Navigate NSFG Website

http://www.cdc.gov/nchs/nsfg/index.htm



Navigate NSFG Website (Continued)





2011-2013 NSFG webpage

NSFG website provides information on each NSFG survey.

Take 2011-2013 NSFG as an example, the website provides:

- Interactive Codebooks and User's guide
- Design and Data Collection Methods
- Variance Estimation Examples
- Questionnaires
- Informed Consent Materials
- The location of downloadable Data Files
- Program Statements for SAS, Stata, and/or SPSS.
- The availability of other Data Files



Questionnaire

NSFG questionnaire has two formats:

1. CAPI-Lite format provides abridged representations of the question wording variants and short descriptions of skip patterns through the interview

```
TIMESMAR

CA-1. (Including your present marriage,) how many times have you been married?

Number

[ CA-2, CA-2b, & CA-2c ARE INTENDED TO OBTAIN NAMES OR INITIALS OF HUSBANDS, ONLY FOR PURPOSES OF LOOPING THROUGH CA SERIES.
[ IF R HAS ANSWERED DK/RF FOR # OF TIMES MARRIED, SHE IS LOOPED ONLY ONCE THROUGH CA SERIES.
```



Questionnaire (Continued)

2. CAPI Reference Questionnaire (CRQ) provides comprehensive specifications of each question that NSFG staff provided to the computer programmers to created the CAPI instrument.

| | Including y | = 1 or 4 (married or separated), ASK: your present marriage, how many times have you been married? ARIT = 2 or 3 (widowed or divorced), ASK: mes have you been married? |
|-------------|--------------------|---|
| | | TER number |
| | (HELP AVAILABLE) | |
| | | er . |
| | | MLYING RANGE: 1 to 95 we want to allow space for only 6 husbands in looping) |
| | FLOW CHECK C-4: | IF MARSTAT = 1 (married) AND 1 LE TIMESMAR LE 6, ASK CA-2 HUSBNAMEX FOR EACH FORMER HUSBAND IF ANY (up to TIMESMAR-1) AND THEN ASK CA-2b HSBVERIF TO VERIFY NAME OF CURRENT HUSBAND. |
| | | ELSE IF 2 LE CA-1 TIMESMAR LE 6 OR (CA-1 TIMESMAR = 1 AND AB-1 MARSTAT NE 1), ASK CA-2 HUSBNAMEX FOR EACH OF THE [TIMESMAR] FORMER (or currently separated) HUSBANDS. |
| | | ELSE IF MARSTAT = 1 (married) AND 1 LE TIMESMAR LE 6 AND CHPNAME = "YOUR HUSBAND", ASK CA-2 HUSBNAMEX FOR EACH FORMER HUSBAND IF ANY (up to TIMESMAR-1) AND THEN GO TO CA-2c CHVERIFY TO FIND OUT NAME OF CURRENT HUSBAND. |
| | | ELSE IF TIMESMAR = DK OR RF, GO TO CA-2c CHVERIFY TO FIND OUT NAME OF CURRENT OR MOST RECENT HUSBAND (BASED ON MARSTAT VALUE). |
| | | ELSE IF TIMESMAR GT 6 AND MARSTAT = 1 (married), ASK HUSBNAMEX FOR HER FIRST 5 FORMER HUSBANDS AND THEN ASK CA- 2c CHVERIFY OR CA-2b HSBVERIF (as appropriate based on CHPNAME value) ABOUT R's CURRENT HUSBAND. (For currently married women with more than 6 marriages, we will only ask about her 1° 5 husbands and her current husband.) |
| | | ELSE IF TIMESMAR GT 6 AND MARSTAT NE 1 (married), ASK |
| | | HUSBNAMEX FOR HER FIRST 6 FORMER HUSBANDS. (For currently |
| | | unmarried women with more than 6 marriages, we will only ask about her I** 6 marriages.) |
| Fan Den Den | nily and nographic | Research |
| | | |

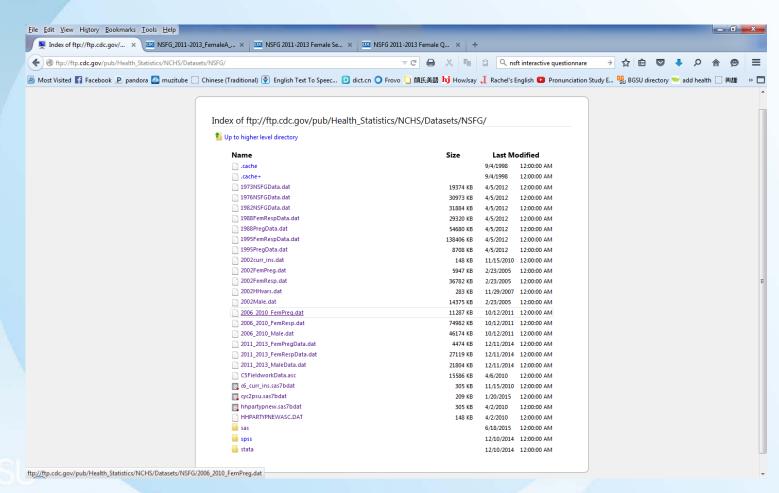
Interactive Codebook





Download Data

NSFG data can be downloaded from the NSFG FTP site: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSFG/





Weighting in NSFG Analyses

NSFG analysis needs to be weighted to adjust for:

- The complex survey design (not a simple random sampling)
- Oversample sub-populations
- Screener and interview non-response
- Estimated characteristics of population from the Census Bureau

The name of the weight variables may change across NSFG cycles.

| Table 3. Weight Variables in 2006-2010 NSFG and 2011-2013 NSFG | | | | | |
|--|------------------------|-----------------------|--|--|--|
| | Variables in 2006-2010 | Variable in 2011-2013 | Variable Description | | |
| | SEST | SEST | The name of the stratum variable | | |
| | SECU | SECU | The name of the cluster variable | | |
| | wgtq1q16 | wgt2011_2013 | Final weight personal weight for the study period | | |
| | wgtq1q8 | | Final weight for years 1 and 2 of data collection (quarters 1 to 8) | | |
| | wgtq9q16 | | Final weight for years 3 and 4 of data collection (quarters 9 to 16) | | |
| | finalwgt30 | | Final weight for the 1st 2 1/2 years of data collection (quarters 1 to 10) | | |
| | wgtq5q16 | | Final weight for years 2, 3, and 4 of data collection (quarters 5 to 16) | | |
| Center for | Family and Demographi | C Research | | | |

Reading NSFG 2011-2013 Data

Stata: Modify the following lines in the Stata command file from the NSFG website.

```
cd "C:\temp\nsfg\nsfg\NCHS_7_11_16\2011_2013" local raw_data "2011_2013_FemPregData.dat" local dict "2011_2013_FemPregSetup" local outfile "2011_2013_FemPregData.dta" . . . . . . . . . . save `outfile', replace
```



Linking A Female Respondent File with A Pregnancy File

```
Sort the female respondent data by ID
use 2011_2013_FemRespSetup.dta, clear
sort caseid
save, replace
 Sort the pregnancy data by ID and pregnancy order
use 2011_2013_FemPregData.dta, clear
sort caseid pregordr
save, replace
   ********
* Merge and save data
use 2011_2013_FemRespSetup.dta, clear
merge 1:m caseid using 2011_2013_FemPregData.dta
save 2011 2013 fem_pregnancy.dta, replace
```

Family and Demographic Research



Pooling Men and Women Data from the Same Time Period

* Select variables from the male file

use 2011_2013_MaleSetup.dta, clear keep caseid rscrninf wgt2011_2013 secu sest save temp1.dta, replace

* Select varaibles from the female file

use 2011_2013_FemRespSetup.dta, clear keep caseid rscrninf wgt2011_2013 secu sest save temp2.dta, replace

* Pool two data sets together

use temp1.dta, clear append using temp2.dta

save 2011_2013_male_female.dta, replace



Using Weight Variables in Analyses

| Table 3. Nine examples of using weight varaibles from the NSFG analyses | | | | | |
|---|---|--|--|--|--|
| 1 | Percent of women using the oral contraceptive pill by age | | | | |
| 2 | Total number of women using the oral contraceptive pill by age | | | | |
| 3 | Mean numbers of children ever born (PARITY), by race and Hispanic origin for women 20-44 years of age | | | | |
| 4 | Linear regression of PARITY on age, race and Hispanic origin, and education for women 20-44 years of age | | | | |
| | | | | | |
| 5 | Percent distribution of pregnancies by wantedness status, by race and Hispanic origin, and age at pregnancy outcome | | | | |
| | Percent of male and female teenagers 15-19 years of age who have ever had sex, by gender and race and Hispanic | | | | |
| 6 | origin (combined male/female files) | | | | |
| 7 | Percent of males 20-44 years of age who have ever fathered a child, by race and Hispanic origin | | | | |
| | Percent of males and females who strongly agree that "a young couple should not live together unless they are | | | | |
| 8 | married" by gender and age | | | | |

Logistic regression for the probability of strongly agreeing that "a young couple should not live together unless they

are married" by age, gender, race and Hispanic origin, and education



9

Conclusions

- Since 2002, NSFG provide data from a nationally representative sample of men and women aged 15-44. The age range has been expanded to 15-49 in the most recent NSFG survey that is currently in the field.
- NSFG provides a unique opportunity to study life events (marriage, cohabitation, and fertility) and attitudes toward marriage and family. New NSFG data should come out this fall and will be great for new research project.
- Users should always use recoded variables, instead of raw variables because recoded variables had been corrected for possible errors and inconsistencies.
- You should always include weight variables in the analyses to take into account NSFG's complex sample design
- NSFG uses a repeated cross-sectional study design and respondents were not followed over time. In addition, female respondents and male respondents in the NSFG were selected from different households and are not couples.

If you have any questions about using NSFG data. Please feel free to contact placeh-Sheng Wu @372-3119 or wuh@bgsu.edu 25

Lisueh-Sheng Wu @372-3119 or wuh@bgsu.edu 25