SAFER • HEALTHIER • PEOPLE™

CDC Monitoring the Nation's Health

Vital and Health Statistics

Series 2, Number 150

June 2010

The 2006–2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey

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

Suggested citation

Lepkowski JM, Mosher WD, Davis KE, Groves RM, Van Hoewyk J.
The 2006–2010 National Survey of Family Growth: Sample design and analysis of a continuous survey. National Center for Health Statistics. Vital Health Stat 2(150). 2010.

Library of Congress Cataloging-in-Publication Data

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402-9328
Printed on acid-free paper.

Vital and Health Statistics

Series 2, Number 150

The 2006–2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey

Data Evaluation and Methods Research

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

Hyattsville, Maryland
June 2010
DHHS Publication No. (PHS) 2010–1350

Contents

Appendixes

Figures

Text Tables

Acknowledgments

This report describes the sample design of the 2006–2010 National Survey of Family Growth (NSFG), which uses a new continuous design, and its implications for statistical research. Subsequent reports in Series 2 will present detailed results on final response rates, final weights, results of imputation, and further guidance for researchers intending to use the NSFG. The survey was designed and conducted by the National Center for Health Statistics (NCHS) and its survey contractor, the Institute for Social Research (ISR), University of Michigan, of Ann Arbor, Michigan. The sampling plan was developed by Robert M. Groves, Steven G. Heeringa, and James M. Lepkowski of ISR, in consultation with William Mosher and Karen Davis of NCHS. The information in this report is based on survey design documentation prepared by staff of ISR and on internal NCHS memoranda.

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

- The Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD)
- The Office of Population Affairs (OPA)
- The CDC's National Center for Health Statistics (NCHS/CDC)
- The CDC's Division of HIV/AIDS Prevention (DHAP/CDC)
- The CDC's Division of Sexually Transmitted Disease Prevention (DSTDP/CDC)
- The CDC's Division of Reproductive Health (DRH/CDC)
- The Children's Bureau of the Administration for Children and Families (ACF)
- The Office of the Assistant Secretary for Planning and Evaluation (OASPE).

NCHS gratefully acknowledges the contributions of these programs and agencies, and all others who assisted in designing and carrying out the NSFG.

The authors of this report gratefully acknowledge the comments of Lester R. Curtin, Ph.D., of NCHS for his peer review of the manuscript. This report was prepared under the general direction of Charles J. Rothwell, Director of the Division of Vital Statistics (DVS), and Stephanie J. Ventura, Chief, of the Reproductive Statistics Branch (RSB). The report was edited by Gail V. Johnson, CDC/NCHM/Division of Creative Services, Writer-Editor Services Branch. Typesetting was done by Zung T. Le and graphics were produced by Odell Eldridge (contractor), NCHS/OD, Office of Information Services, Information Design and Publishing Branch.

Abstract

Objective

The National Survey of Family Growth (NSFG) collects data on pregnancy, childbearing, men's and women's health, and parenting from a national sample of women and men 15–44 years of age in the United States. This report describes the sample design for the NSFG's new continuous design and the effects of that design on weighting and variance estimation procedures. A working knowledge of this information is important for researchers who wish to use the data.

Two data files are being released—the first covering 2.5 years (30 months) of data collection and the second after all data have been collected. This report is being released with the first data file. A later report in this Series will include specific results of the weighting, imputation, and variance estimation.

Methods

The NSFG's new design is based on an independent, national probability sample of women and men 15–44 years of age. Fieldwork was carried out by the University of Michigan's Institute for Social Research (ISR) under a contract with the National Center for Health Statistics (NCHS). In-person, face-to-face interviews were conducted by professional female interviewers using laptop computers.

Results

Analysis of NSFG data requires the use of sampling weights and estimation of sampling errors that account for the complex sample design and estimation features of the survey. Sampling weights are provided on the data files. The rate of missing data in the survey is generally low. However, missing data were imputed for about 600 key variables (called "recodes") that are used for most analyses of the survey. Imputation was accomplished using a multiple regression procedure with software called IVEware, available from the University of Michigan website.

Keywords: survey methodology • imputation • variance estimation • continuous interviewing

The 2006–2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey

by James M. Lepkowski, Ph.D., Institute for Social Research, University of Michigan; William D. Mosher, Ph.D., and Karen E. Davis, M.A., National Center for Health Statistics; Robert M. Groves, Ph.D., and John Van Hoewyk, Ph.D., Institute for Social Research, University of Michigan

Executive Summary

The National Survey of Family Growth (NSFG) obtains detailed information on factors affecting childbearing, marriage, and parenthood from a national probability sample of women and men 15–44 years of age. This Series 2 report describes the procedures used to select the sample, develop the sampling weights, impute missing data, and estimate sampling errors. This information should be useful for those who intend to do statistical research with NSFG data and for survey methodologists who want to compare their procedures to those used in the NSFG.

This report is a significant departure from previous NSFG Series 2 reports in two ways. First, previous NSFG Series 2 reports have been released 3–4 years after data collection ended, which meant that many data analysts had to do their initial analyses before the Series 2 report was available. This report, in contrast, is being released to coincide with the first release of data from the continuous NSFG, so that NSFG data users can use the information when they are ready to do their initial analyses. Second, in continuous interviewing, specific results for each public-use data file will be released as soon as they are available on the NSFG website and in subsequent Series 2 reports. This procedure should deliver more information into the hands of data users in a timely way than in past NSFG cycles. This report describes the sample design and related topics for the first 4 years of the continuous NSFG.

The NSFG is designed and administered by the National Center for Health Statistics (NCHS), an agency of the U.S. Department of Health and Human Services, in response to Section 306 of the Public Health Service Act, which directs NCHS to "collect statistics on . . . family formation, growth, and dissolution," as well as "determinants of health" and "utilization of health care." Accordingly, the purpose of the survey is to produce reliable national statistics on:

- 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
- What women and men do to raise their children
- Men's and women's attitudes about sex, childbearing, and marriage

The 2006–2010 NSFG was conducted by the University of Michigan's Institute for Social Research (ISR) under a contract with NCHS. The 2006–2010 NSFG was the first time the NSFG was fielded using a **continuous design**. A "continuous design" means that NSFG interviewing will be done every year as long as funding and other circumstances permit. The fieldwork plan for the first continuous interviewing sample called for interviewing to be done in 4 years. Interviewing for the 2006–2010 survey began on or about July 1, 2006. Interviews were conducted with a national probability sample of women and men 15–44 years of age living in households in the United States. The interviews were administered in person by trained female interviewers using laptop or notebook computers, a procedure called computer-assisted personal interviewing (CAPI). The interviews were designed to average approximately 80 minutes for women and 60 minutes for men.

The 2006–2010 NSFG was based on a sampling plan that was intended to provide larger and more cost-effective samples than ever before in the NSFG's history. The 2006–2010 national sample was drawn from 110 major areas, or primary sampling units (PSUs), divided into four national subsamples. Each of the four subsamples was worked for 1 year, so the entire 110–PSU design could be completed in a 4-year period. The entire 4-year data file is expected to yield at least 5,000 interviews per year, or 20,000 interviews in 110 PSUs for the full 4-year time period (mid-2006 to 2010)—the largest sample in the NSFG's history.

Black, Hispanic, teenage, and female respondents were sampled at higher rates than others. Sampling weights were used to compensate for the different sampling rates of these various groups, and for different nonresponse rates. Sampling errors were estimated using software that takes the weights and the stratified cluster sample design features into account. Such software is now widely available in such packages as SAS, Stata, SPSS, SUDAAN, and others.

In addition, for key variables, referred to in this and other NSFG reports as ''recodes,'' item-missing values have been replaced in the data file by predicted or imputed values. The imputed values in the approximately 600 ''recode'' variables were identified with a companion variable (or ''imputation flag'') that indicates whether the value for a particular case was imputed or reported.

The rest of this report describes how the sample was designed and selected, how sampling weights are computed and adjusted to compensate for the different sampling rates and other factors, and how missing data are imputed for selected recodes. This report concludes with a section on ''NSFG Continuous Design: A Guide for Analysts,'' which includes some recommendations for those wishing to do research with the NSFG. The section includes 10 Frequently Asked Questions (FAQ's) to clarify these practical issues for analysts further.

The first data release includes over 13,000 interviews conducted from about July 1, 2006, through December 2008. It is not expected that all readers will need to read every section of this report. As a result, the report is designed to describe the 2006–2010 NSFG sample design at three levels of detail:

- First, this Executive Summary and the ''Guide for Analysts'' are for readers seeking a general, nontechnical understanding of the survey procedures.
- Second, summaries for each major section of the text are included to provide somewhat more information on each design feature.
- Third, in the rest of the report, the full technical details are provided for the interested reader. Thus, those who read the entire report will find some topics described first in less detail in the section summaries and then in full detail.

Development of the National Survey of Family Growth

The NSFG was established at the National Center for Health Statistics (NCHS) in 1971. Cycle 1 was conducted in 1973 (1). Before then, smaller national surveys of married women were conducted by private organizations in 1955 and 1960 (2,3). In 1965 and 1970, they were conducted by university researchers with federal funding (4,5).

As shown in Figure 1, the NSFG has been conducted seven times since 1973 by NCHS. A ''cycle'' consisted of planning, pretest, fieldwork, data processing, file preparation, and documentation for a single survey, but the year given is the year the interviews, or most of the interviews, were done. Cycle 1 interviewing was conducted in 1973, when 9,797 women 15–44 years of age were interviewed, making Cycle 1 the largest sample at that time for a U.S. national fertility survey. Cycle 1 in 1973 and Cycle 2 in 1976 were restricted to women who were currently or formerly married and focused primarily on pregnancy history, contraceptive use, birth intentions, marriage histories, and a variety of social and economic characteristics (1,6). Cycles 1 and 2 interviews were conducted in either English or Spanish, at the respondent's choice. This practice has continued through all NSFG surveys.

NSFG Cycle 3 in 1982 expanded the sampling frame to include all women aged 15–44 regardless of marital status, making it possible to study the contraceptive use, sexual activity, and use of family planning services of unmarried women and teenagers as well as the married population (7). Cycle 3 sampled teenage females (15–19 years of age) at a higher rate to gain information relevant to growing concern over teenage pregnancy.

NSFG Cycle 4, fielded in 1988, responded to important fertility-related questions of the day, including detailed questions on cohabitation, adoption, and sexually transmitted diseases. Specifically, new questions covering respondents' knowledge of chlamydia, genital herpes, and AIDS-related knowledge and behavior were introduced in Cycle 4 (8).

Several changes were made in NSFG Cycle 5, conducted in 1995, in response to recommendations that the NSFG should increase the number and depth of measures used to predict fertility-related variables. These changes included the first oversample of Hispanic women, and converting the survey from paper-and-pencil

Cycle	Year	Survey contractor	Scope or population covered	Number of interviews	Source of sample	Over-samples	Average length in minutes	Incentive payment
1	1973	NORC	Ever-married women 15–44	9,797	Independent 101 PSUs	Black women	60	No
2	1976	Westat	Ever-married women 15–44	8,611	Independent 79 PSUs	Black women	60	No
3	1982	Westat	All women 15–44	7,969	Independent 79 PSUs	Black women teens	60	No
4	1988	Westat	All women 15–44	8,450	NHIS 156 PSUs	Black women	70	No
5	1995	RTI	All women 15–44	10,847	NHIS 198 PSUs	Black women Hispanic women	100	$20
6	2002	University of Michigan ISR	Men 15–44 Women 15–44	12,571	Independent 121 PSUs	Blacks Hispanics teens	Men=60 Women=85	$40
Continuous	2006–2010	University of Michigan ISR	Men 15–44 Women 15–44	5,000 per year	Independent 110 PSUs in 4 years	Blacks Hispanics teens	Men=60 Women=80	$40

NOTE: PSU is primary sampling unit. NHIS is National Health Interview Survey.

Figure 1. History of the National Survey of Family Growth

interviewing to computer-assisted personal interviewing (CAPI) in an effort to improve the consistency and quality of the data. In addition, several event histories were introduced to increase the analytic usefulness of the birth and pregnancy, contraception, and marital histories the NSFG had always collected. Those histories included an education history, a work history, and a cohabitation history. A life history calendar was added to help organize all the event history information. A file of contextual data was also created, allowing researchers to examine the ways in which characteristics of the place of residence—census tract, local, or state—influence behaviors. Information was also collected on respondents' sexual partners, the wantedness of pregnancies, the consistency of contraceptive use, the circumstances under which first intercourse occurred, and the use of family planning services. On the technical side, CAPI on laptop computers replaced the paper and pencil questionnaire, improving both the quality and timeliness of the data. Another innovation was the use of audio computer-assisted self interviewing (ACASI), in which respondents used laptops to hear and read the most sensitive questions and enter answers by themselves (9,10).

In Cycle 6, conducted in 2002, the sample was expanded to include males 15–44 years of age, with a new questionnaire specially designed for them. The male questionnaire and other changes were informed by a set of reports from survey experts as well as by a large pilot study conducted in 2001. The extensive planning and research for Cycle 6 have been described elsewhere (11) (see http://www.cdc.gov/nchs/data/series/sr_01/sr01_042.pdf).

The experience of the 2002 NSFG was used to inform the design of the continuous NSFG, where a ''continuous design'' simply means that NSFG interviewing will be done every year indefinitely, as long as funding and other circumstances permit. The fieldwork plan for the first continuous interviewing sample was to complete interviewing in 4 years. After that, another sample would be drawn for the next few years of interviewing. The continuous interviewing design was an innovation in the NSFG series motivated by the need to produce larger sample sizes with a constant budget, in the face of increasing uncertainty in the performance of large, complex face-to-face surveys in the United States. These uncertainties arise because of unknown eligibility rates in samples of U.S. addresses, unknown contact and cooperation rates, and various inefficiencies in the staffing organization of large one-time field effort designs. The 2002 NSFG utilized over 250 interviewers over a 12-month period. This expansion raised several challenges: quality control with large sets of interviewers working for only a few months is difficult; the survey costs are increased by the inevitable learning curves that new interviewers experience in the first months; attrition among interviewers is more likely with small workloads; and interviewer workloads can be inefficient in size, which reduces average interviewer productivity.

To produce more predictable results and to better control costs, NSFG continuous interviewing used about 40 interviewers working consistently over the year with workloads designed to maximize their productivity. The sample design of continuous interviewing (described in the following text) attempted to deal with each of these challenges to control costs and increase quality.

The NSFG provides data needed by several federal programs in addition to NCHS, including:

- Programs of the Office of Population Affairs (OPA). This office is concerned with teenage sexual activity and pregnancy, and the use of Title X Family Planning Services. OPA is the lead agency in charge of the Healthy People 2010 and 2020 objectives on Family Planning, most of which are based on NSFG data.
- The Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD) uses the data to inform and shape its extramural (grant) research programs related to fertility and reproductive health in the United States and to provide a data resource for private and university-based researchers.
- The Centers for Disease Control and Prevention's (CDC) Division of HIV/AIDS Prevention (DHAP/CDC) and the Division of Sexually Transmitted Disease Prevention (DSTDP/CDC) need reliable information on the sexual and drug-use-related behaviors that increase the risk of HIV and STD transmission.
- Other programs also use and support the NSFG, including the Office of the Assistant Secretary for Planning and Evaluation (OASPE)—for a wide range of data including data on men and fatherhood; the Children's Bureau of the Office of Planning Research, and Evaluation of the Administration for Children and Families (ACF)—for data on adoption and related issues; and the CDC's Division of Reproductive Health (DRH/CDC)—for data on teenage pregnancy and unintended pregnancy, among other issues.

Design Specifications

The NSFG's 2006–2010 sample design was based on the following objectives:

- The target population for the continuous NSFG was to be the household population of men and women aged 15–44 in households in the 50 states and the District of Columbia.
- Screening interviews were to be conducted in each sampled household, to determine if anyone 15–44 years of age lived there and if so, to select one person from the household for the NSFG interview.
- The selection was to be random but to sample certain subgroups at higher rates. Those groups included teenagers (15–19 years of age), Hispanic men and women, and non-Hispanic black men and women.
- Data collection was to be conducted only by in-person, face-to-face interviewing, with the respondent's privacy and confidentiality ensured.
- Questionnaires and interviews were to be available in English and Spanish.
- Interviews were expected to last an average of 80 minutes for women and 60 minutes for men.
- Given the sensitivity of the interview content, all interviewers were to be female. That is, men and women were to be interviewed by a female interviewer.
- Data collection was to be completed using CAPI. One section of the questionnaire was to be administered using ACASI, in which the respondent would listen to a pre-recorded audio reading of the questions with a headset connected to the laptop. The questions were also to be displayed on the computer screen. Survey instruments were to be programmed on laptop computers meeting specified requirements.
- The available funds appeared to allow a sample of approximately 4,400–5,000 interviews per year, yielding 17,000–20,000 interviews over a 4-year data collection period. Of these, about 45 percent were to be males and 55 percent females; about 20 percent teenagers 15–19 years of age; about 20 percent Hispanic; and about 20 percent black or African American.
- The following sample sizes and design effects were expected (Table A).
- The sample was to be designed so that any number of individual years could be combined to form a nationally representative sample for analysis. Practical considerations, such as sample size and design effects, however, made it likely that more than 1 year of data would be needed for most analytical purposes.
- Signed informed consent was required for every selected eligible respondent 15–44 years of age. Minors 15–17 years of age were required to have the signed consent of a parent before being asked for their own signed assent.
- Interviewers were to ask questions about fertility, contraceptive use, sources and types of family planning services, and maternal and child health using structured questions similar to those used in the 2002

Table A. Sample sizes and design effects by selected characteristics

Characteristic	Sample N	Design effect
Total	17,600–20,000	3.9
15–19	3,500–4,000	1.5
Race		
Hispanic	3,500–4,000	1.6
Black or African American	3,500–4,000	1.6
Sex		
Male	7,900–9,000	2.2
Female	9,700–11,000	2.8

(Cycle 6) NSFG, and read by an interviewer (in the CAPI part of the interview) or heard on a headset (in the ACASI part of the interview).

- The contractor, in cooperation with NCHS, was to design and implement procedures for measuring and controlling the quality of data collection and data preparation procedures, including verification of a sample of interviews.

Sample Design

Overview of Sample Design

The first step in the sampling procedure was to select a national sample of 110 primary sampling units (PSUs). PSUs are counties or groups of adjacent counties. This national sample was divided into four parts, each of which was a nationally representative sample. The fieldwork plan envisioned working the sample over a 4-year period, so one of these four *national quarter samples* was used each year. As this report was written, an average of about 5,500 men and women were being interviewed each year.

The PSUs used in each annual quarter sample consisted of the eight largest metropolitan areas in the United States (which become eight individual PSUs in the sample and remain in the sample each year) and 25 smaller metropolitan and nonmetropolitan PSUs that change each year. (In the first year, two additional PSUs were included in the sample for purposes of more detailed representation of the United States, for a sample of 35 PSUs in the first year, so the number of PSUs was 35 the first year, 33 the second year, 33 the third year, and 33 the fourth year, for a total of 110.) The national sample of 110 PSUs, which is used over a 4-year period, includes PSUs in most states.

From each PSU, secondary units, called segments, were selected. Segments are, roughly, neighborhoods or groups of adjacent blocks. In each selected segment, one of two procedures has been used to obtain a housing unit sample. In most segments, segment addresses were obtained from a commercial source or from prior Cycle 6 listed addresses, visited by interviewers to check and correct list accuracy, and selected by central office staff at ISR for screening and interviewing. In other segments, interviewers visited the segment and listed all housing units (''from scratch''). A sample of addresses is selected from the interviewer list by ISR central office staff. More details on the stages of sample design can be found in the ''Detailed Sample Design'' section.

The sample housing units were then contacted, and a ''screener'' interview attempted, in which the persons living at that address (including persons living away from the household in a college dormitory, sorority, or fraternity) were listed. If more than one eligible person 15–44 years of age was living at the address, one person was randomly selected and asked to do an interview.

The NSFG sample design consisted of five stages of selection to choose eligible sample persons. Women, teens 15–19 years of age, and black and Hispanic persons are selected at higher rates, yielding an oversample of such persons.

The 2006–2010 NSFG sample design started with the same national sample of **PSUs** used in the 2002 (''Cycle 6'') NSFG national sample design. The same sample of PSUs was used in 2002 and in 2006–2010 because it reduced the cost of sample selection for the NSFG, and because gains in precision could be achieved when comparing Cycle 6 to continuous NSFG findings when the same PSUs were used.

The PSU selection in the 2006–2010 NSFG began with using the U.S. Census Bureau division of the entire land area of the United States into 3,141 counties and county-equivalent units. These units were either used individually or grouped to form 2,402 PSUs. In most cases, a PSU was a single county. But for large metropolitan areas, counties were grouped together by the U.S. Census Bureau to form what the Bureau calls metropolitan statistical areas. In the 2002 and the 2006–2010 NSFG sample design, then, a county in a metropolitan area was grouped with other counties in the same metropolitan area to form a PSU. These 2,402 PSUs were based on the 2000 Census definitions of the units.

Following the creation of the PSUs, a process called stratification was used to partition the PSUs into three major groups or strata: 28 large metropolitan areas, 290 other metropolitan areas, and 2,084 nonmetropolitan areas. The 28 large metropolitan areas are referred to as self-representing (SR) areas. SR areas are those that have such large populations that a national sample of the size used for continuous NSFG virtually required that they be represented. As such ''certainty'' selections, the sample from each of these areas represents only those areas. That is, the sample from these 28 PSUs represents only the population of that area. Hence, in the sampling literature, these types of units are referred to as representing only themselves, or ''self-representing.''

The remaining 2,374 PSUs are called non-self-representing (NSR) areas. A sample of the NSR PSUs was selected so that each sample PSU represented itself and other NSR PSUs of a similar nature. In order to make the representation more complete, the NSR PSUs were further grouped by geography and population size into 82 sets or strata. Each NSR stratum had two or more PSUs, and some strata had more than 100 PSUs. The number of PSUs in a stratum varied because the strata were created to have approximately equal 2000 Census population across the PSUs.

In total, the 2006–2010 NSFG consisted of 2,402 PSUs grouped into 110 strata. Twenty-eight of the strata contained a single PSU—the SR strata (with one PSU each). The remaining 82 strata contained two or more PSUs each, one of which was selected. (In other words, in the SR strata the single PSU was certain to be selected). In the NSR strata, one PSU was selected from those within the stratum to represent the entire stratum.

The selection of the 82 NSR PSUs from the NSR strata was, for purposes of sample size control, selected with a technique called ''probability proportional to size.'' This technique uses information about the number of *housing units or number of persons* to select PSUs. In particular, PSUs with larger populations have higher chances of selection, and those with smaller populations have lower chances. Within a PSU, the sample is selected with chances that are inversely proportionate to those used to select the PSU. The PSU and within PSU chances are carefully calibrated in the sample selection to eliminate any over- or under-representation of the population. This result is referred to as equal chance selection, or a self-weighting sample.

A single PSU was selected from each of the 110 strata in the 2006–2010 NSFG national sample. The sample of 110 PSUs was then divided into four fully representative national samples to allow new samples to be introduced each data collection year, to reduce interviewer and field staff workload; to control costs more effectively; and to allow accumulation of sample and data across years to yield larger sample sizes across longer time periods.

Each annual *national quarter sample* consists of the eight largest SR metropolitan areas (referred to as the ''super eight PSUs'') that were in the sample in each of the 4 years; 5 of the 20 SR metropolitan areas, and 20 (or 22) NSR metropolitan and nonmetropolitan PSUs from the 82 NSR PSUs. One national quarter sample was chosen at random and without replacement for each year of interviewing. Interviewing began effectively July 1, 2006 (a few interviews were conducted in late June).

In the **second stage of selection**, census blocks were stratified into four **domains** within each PSU, and the housing units on those blocks were listed. The four domains were defined to obtain oversamples of black and Hispanic persons. They were:

1. Nonminority
2. More than 10 percent black persons but less than 10 percent Hispanic persons
3. More than 10 percent Hispanic persons but less than 10 percent black persons
4. More than 10 percent black persons and more than 10 percent Hispanic persons.

To select the national sample for each data collection year, blocks were chosen within each domain with probabilities proportionate to an estimated number of Census 2000 occupied housing units. Small blocks with only a few occupied housing units were linked with other blocks to form ''segments'' large enough to support efficient data collection workloads. In NSR PSUs, 12 segments were selected for a single year of data collection. Larger numbers of segments were selected in most of the SR PSUs, where a larger number was needed from larger PSUs. Within both SR and NSR PSUs, the segments selected for a given data collection year were randomly divided into four sets. One set was released and completed in each calendar quarter (beginning July 1, October 1, January 1, and April 1).

The housing units on the selected blocks were listed. (The process of obtaining lists of addresses is described in detail in the section ''Field listing of segments.'') Then the housing unit lists were uploaded to ISR and checked by central office staff for consistency and accuracy.

The third stage of selection chooses **housing units** from the list of addresses available in each sample segment. Housing units in census blocks in which more than 10 percent of the population was black or Hispanic persons (domains 2, 3, and 4) are selected at higher rates to increase the number of black and Hispanic persons in the sample. Data for selected housing units are downloaded to the interviewers' laptop computers for the next stage of sampling, and interviewing.

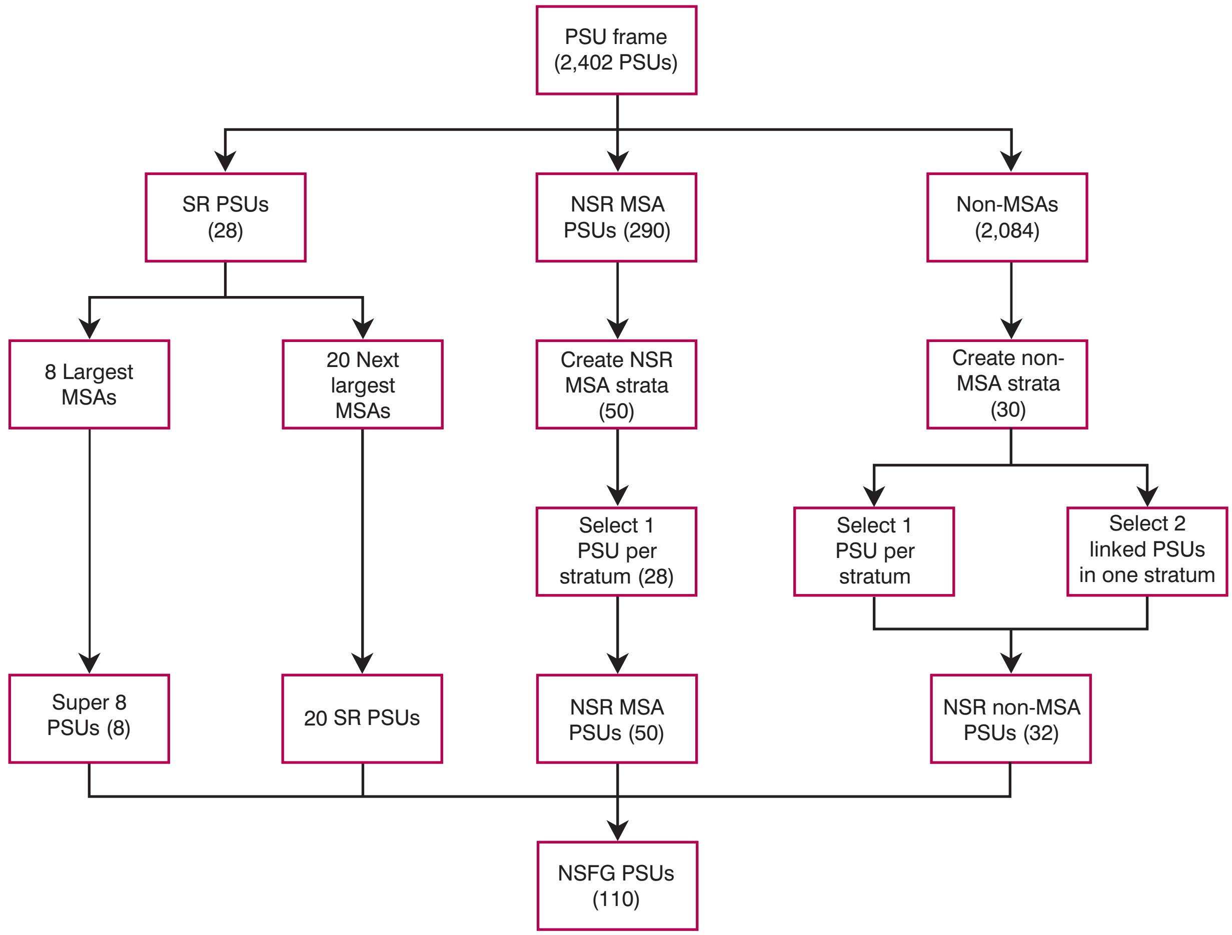
The fourth stage of sampling is the selection of eligible **persons** within sample households. Interviewers visit housing units selected in the third stage, and when the housing unit is found to be occupied, attempt to list all persons living there. One eligible person is chosen randomly in every household containing one or more eligible persons. The within-household selection used measures of size that are pre-assigned in the sample design process. These measures of size vary by gender, age group, and race and ethnicity. Teenagers 15–19 years of age and females receive larger measures of size, and thereby increased chances of selection within the household, in order to yield larger sample sizes required for the target sample sizes for these groups. Persons within households are selected with a probability proportionate to a measure of size. An example of this unequal probability sampling procedure is described in the section on ''Detailed Sample Design.''

Persons living away from home in college or university dormitories, sororities, and fraternities are considered to be part of their parent's household and listed in them. If a college student is chosen, she or he is transferred to an interviewer working in the nearest PSU to the college or university for interviewing. In many cases, college student interviews are obtained at the sample household at times when selected students return home for summer or holidays.

The fifth stage in sample selection occurs in each 12-week ''quarter'' of interviewing: the selection of the ''double sample'' (because it is a sample of a sample). After 10 of the 12 weeks of data collection in each 12-week quarter, a set of selected housing units has not been successfully screened or, if successfully screened, the sampled person has not been interviewed yet. Nearly all such housing units have been contacted, but not all have been screened. To increase representation of these types of housing units in the final sample, a sample of about one-third of these remaining addresses was selected for interviewing at the end of the 10th week. For the last 2 weeks of the quarter, interviewer assignments were reduced to housing units in the second phase sample only. Interviewers concentrated on a smaller number of housing units for the final 2 weeks of data collection.

Detailed Sample Design

The following sections describe the sampling process in the 2006–2010 NSFG in more depth. The steps are shown graphically in Figures 2 and 3.



NOTE: PSU = primary sampling unit; MSA = metropolitan statistical area; SR = self-representing; NSR = non-self-representing; and super 8 = 8 largest PSUs.

Figure 2. National Survey of Family Growth 2006–2010: selection of primary sampling units

First stage: Selection of primary sampling units

Sampling frame

The U.S. Census Bureau in 2000 divided the land area of the United States into 3,141 counties and county equivalents (Louisiana parishes; Alaska boroughs and census areas; independent cities in Maryland, Virginia, Missouri, and Nevada; and the District of Columbia). The ISR grouped these counties into 2,402 PSUs. These PSUs served as the sampling frame for the first stage selection in the 2006–2010 NSFG.

The U.S. Census Bureau also designates counties as part of metropolitan areas. A total of 1,057 counties were grouped into 318 MSAs consisting of two or more geographically adjacent counties. The ISR PSU sample designated these MSAs as separate metropolitan sampling units. The remaining 2,084 nonmetropolitan counties are treated as individual sampling units. This yields a set of 2,402 PSUs in the ISR sample.

The same set of PSUs was used for the 2002 NSFG and the 2006–2010 NSFG. This was done to increase quality and reduce costs of data collection by employing interviewers in both the 2002 and the 2006–2010 NSFG, and using housing unit lists that had been prepared for the third stage of selection in 2002 that had not been selected in the 2002 NSFG in the 2006–2010 NSFG. In addition, the overlap of PSUs between the 2002 and 2006–2010 NSFG allows gains in precision when comparing results from the two surveys.

The PSU selection for the 2002 NSFG thus serves as the PSU sample for the 2006–2010 NSFG. One potential disadvantage of this approach concerns the timing of PSU selection for the 2002 survey. The PSUs for the 2002 NSFG had to be selected in 2001, before detailed 2002 data collection planning began. ISR staff preferred to use 2000 census counts for occupied housing units in the selection of the PSUs. Such counts were not available at the time of the selection. Rather than delay the start of 2002 data collection, 1990 census counts were used to select the PSUs. Subsequent investigation indicated that the use of the more out-of-date counts in the selection led only to minor losses in precision (and no bias) in estimates from the 2002 NSFG. Thus, the importance of timely data collection necessitated the use of 1990 census counts in the 2002 NSFG. These counts were the basis as well for the PSUs in the 2006–2010 survey. There is a minor loss in precision (or increase in variance) for both surveys because there were changes in the U.S. population, particularly for the Hispanic population, between 1990 and 2006. However, these losses in precision were compensated in part through the stratification and allocation of segments across domains within PSUs (domains were defined by percent black or Hispanic; see ''Second stage: Selection of segments''), where domains are defined by the proportion of the segment's population that is Hispanic or black. The allocation within PSUs was used to obtain larger samples of Hispanics, and improved the efficiency of the stratification of PSUs by using more up-to-date measures of size.

The PSUs varied substantially in size, from as few as 25,000 to as many as 3.2 million 1990 occupied housing units. The 28 largest MSAs each had more than 600,000 occupied housing units, and were so large that they were chosen with certainty for the sample. These 28 MSAs constituted 28 strata in the ISR national PSU sample. Each contained exactly one PSU, and that PSU was ''selected'' into the sample with a probability of 1. Since these PSUs are the only PSUs in each of their strata, they represent only themselves, and are referred to as SR units.

The remaining 2,374 PSUs were designated as NSR because a sample of 82 PSUs was to be selected from among them. Thus, each one of the 82 NSR PSUs would represent itself and other PSUs from the same stratum, in the national sample.

The 2,374 NSR PSUs consisted of 290 metropolitan and 2,084 nonmetropolitan PSUs. These PSUs were grouped into the nine census geographic divisions defined by the U.S. Census Bureau state groups. Within divisions, PSUs were further grouped by ISR into 52 MSA and 30 non-MSA strata on the basis of population size and geography. For example, in the East North Central division, MSA PSUs were grouped into 10 strata, ranging from large MSAs in one state to small MSAs in several states. The remaining non-MSA PSUs (all individual counties) were then grouped into four geographically contiguous sets consisting of 60–80 PSUs each. From the 82 such MSA and non-MSA strata, a single PSU was selected. **Figure 2** summarizes the national PSU sample selection.

The 110 PSUs selected for the 2006–2010 NSFG national sample were for purposes of identification grouped into four types:

1. The eight largest metropolitan areas among the 28 SR strata.
2. The remaining 20 largest SR metropolitan areas.
3. Fifty-two NSR but also metropolitan areas.
4. Thirty NSR but nonmetropolitan areas.

This grouping was then used to divide the sample of 110 PSUs into four fully representative national samples for the 2006–2010 NSFG. The four annual national samples allow new samples to be introduced in each of 4 consecutive years of data collection. The smaller national samples provided opportunity to more effectively monitor field data collection and cost, and operate with a smaller central office staff. These four national samples have allowed ISR to make changes, as required, to data collection, survey questions, and other design features once each year, and have nationally representative samples that could be combined across years. The national samples also can be accumulated across years to yield larger sample sizes across longer multiyear time periods.

Each annual *national quarter sample* consists of:

- All eight of the largest SR metropolitan areas, referred to as the ''super eight PSUs,'' that were, because of size, always in the sample.
- Five of the remaining 20 SR metropolitan areas selected carefully to represent the full set of 20 in each year.
- Twenty (or 22, in the first year) NSR metropolitan and nonmetropolitan areas selected to represent the full set of 82 NSR PSUs in each year.

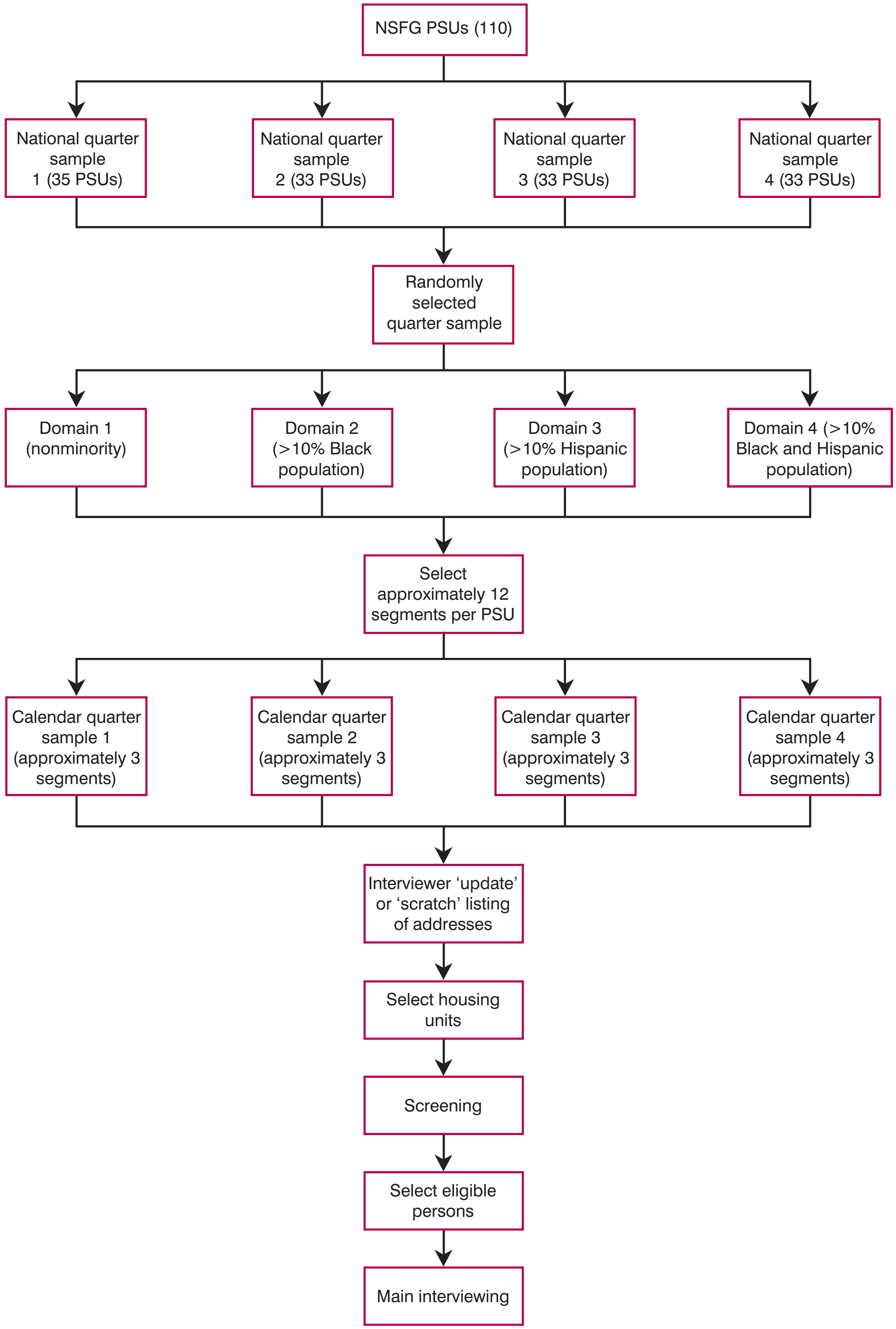
One national quarter sample was chosen at random for the first year of interviewing, which began about July 1, 2006. A second annual (quarter) sample was selected, without replacement, for the year beginning July 1, 2007, a third for the year beginning July 1, 2008, and a fourth for the year beginning July 1, 2009.

Second stage: Selection of segments

Stratification of blocks

The second-stage sampling units were census blocks, or linked combinations of census blocks that had sufficient numbers of households to sustain efficient survey data collection. For each of the 110 PSUs in the NSFG sample, a list of blocks was obtained from 2000 census data (which were available for the 2006–2010 NSFG and for the 2002 NSFG sample). Census blocks have varying numbers of occupied housing units—some blocks having few, if any, occupied housing units. These small blocks pose problems for efficient screening and interviewing. In order to obtain units of sufficient size for efficient data collection, larger blocks or smaller blocks linked to one another to comprise larger units were designated in each PSU. These units are called segments. The minimum size for a segment was 75 occupied housing units in metropolitan areas or 50 in nonmetropolitan areas. Linked blocks were always within the same block group boundaries to reduce the chance that geographically widespread blocks were joined together. The segments were the sampling units for the second stage of selection.

Segments were grouped within each PSU into four strata or domains based on the percentage of black or Hispanic occupied housing units in the segment (see Figure 3). Domain 1, the nonminority stratum, consisted of those segments with less than 10 percent black or 10 percent Hispanic persons in the 2000 census. Domain 2 contained segments with at least 10 percent black occupied housing units, but less than



NOTE: PSU is primary sampling unit.

Figure 3. National Survey of Family Growth 2006–2010: sample allocation summary

10 percent Hispanic; domain 3 contained segments with at least 10 percent Hispanic but less than 10 percent black occupied housing units; and domain 4 contained segments with at least 10 percent black and at least 10 percent Hispanic occupied housing units. The race and ethnicity prevalence rates were computed at the census block group level to reduce classification of geographically contiguous census blocks into different domains.

Black and Hispanic persons were to be oversampled among the final sample of persons in the 15–44 years of age range. Probabilities of selection of segments were varied across domains to increase the chances, and the sample sizes, of black and Hispanic persons in the sample. Higher rates were used for domains 2, 3, and 4, yielding a potential increase in sampling variance, or loss of effective sample size, through weighting to compensate for the unequal sampling rates.

After simulation of design alternatives and examination of sample sizes in target groups and effective sample sizes, sampling rates for domains 2, 3, and 4 were set 200–250 percent higher than the rate for domain 1.

In the 2006–2010 NSFG, exactly 12 segments were selected in 80 of the nonmetropolitan NSR sample PSUs. In the remaining two PSUs, slightly fewer segments were selected, based on the size of these PSUs. The 28 largest PSUs received an allocation of segments that was proportionate to size, with the smallest receiving approximately 12 segments and the largest more than twice as many. (The allocation of the number of segments within a PSU and within a domain is described in the following text.)

Within a PSU, one-quarter of the segments allocated to each PSU in the yearly sample were selected in each 12-week data collection quarter. If there were exactly 12 segments in each of the 33 PSUs in a yearly sample, there would be 3 segments in each PSU in each quarter, or an expected 99 segments in a calendar quarter sample. However, since the eight largest and the five metropolitan SR PSUs had on average more than 12 segments because of their larger populations, the number of segments in a calendar quarter is approximately 110 segments. Over the entire year, approximately 440 segments were in the 2006–2010 NSFG sample.

Over the 4 years, some ''super 8'' segments that have large numbers of housing units in a block within the segment were selected more than once across the years. Across the 2006–2010 period, somewhat less than 4 × 440 = 1,760 segments are being selected. The final number of segments was approximately 1,600 over the 4-year data collection period.

Segment domain allocation

The number of segments to be selected from each domain is determined prior to the national quarter sample release each data collection year. The number of segments in a domain in a PSU depends on the distribution of occupied housing units in the PSU across the ''super 8,'' other SR metropolitan PSUs, and the NSR metropolitan and nonmetropolitan PSUs. It also depends on an allocation for the stratum from which the PSU was selected that is proportionate to the number of occupied housing units in the stratum in each segment domain.

For example, suppose a given NSR PSU *a* is in stratum *h*, and that the proportion of occupied housing units in the stratum across domains 1, 2, 3, and 4 is 0.75, 0.15, 0.05, and 0.05, respectively. If the 12 segments to be selected from stratum *h* (and PSU*a*) were allocated proportionately to these sizes, the allocation would be 9, 1.8, 0.6, and 0.6. However, the sample from domains 2, 3, and 4 must be selected at a rate that is, say, 1.5 times larger than that from domain 1 in order to achieve target sample sizes by race and ethnicity. The allocation must thus be adjusted to account for the differences in the relative sampling rates across the domains. The allocation must thus be made proportionate to occupied housing unit proportions and a domain sampling rate. In this illustration then, the allocation becomes **9 × 1 = 9; 1.8 × 2.5 = 4.5; 0.6 × 2.5 = 1.5;** *and* **0.6 × 2.5 = 1.5,** respectively, across the four segment domains. The segment counts of 9, 4.5, 1.5, and 1.5 are now in the correct proportion to achieve the expected sample distribution, but they total 16.5 instead of the desired 12. The segment domain counts are then adjusted to sum to 12, or **9 × (12/16.5) = 6.54; 4.5 × (12/16.5) = 3.27; 1.5 × (12/16.5) = 1.09;** *and* **1.5 × (12/16.5) = 1.09,** respectively.

These final noninteger values are rounded to integer values across PSUs to maintain the distribution required for target sample sizes. In this illustration, 7, 3, 1, and 1 are to be selected from domains 1, 2, 3, and 4.

In some PSUs, there was not a sufficient number of segments available to support the target sample size in a particular domain, because the PSU in the sample from the stratum had a lower proportion of domain population than did all the PSUs in the stratum. In most cases when this occurred, the number of allocated segments in a PSU differed from the target number by one. This allocation was adjusted in a PSU in another stratum from the same region and of the same size (metropolitan or nonmetropolitan). The allocation for a PSU with insufficient numbers of segments in a given domain was decreased by one in the domain, while the allocation in the PSU from a similar region and size stratum was increased by one in the same domain.

Selection of segments

Once blocks are linked and segment sample sizes determined for a PSU, a sample of segments is selected from each PSU. Within a PSU domain, census blocks are ordered geographically by tract, block group, and block number. Segments were selected within a PSU separately for each domain, systematically with probabilities proportional to the relative number of 2000 census occupied housing units in each segment (see the following text), a method used widely in survey samples to select ''clusters'' of unequal size (see, for example, Chapter 7 in reference 12).

Since the same PSUs were being used for the 2006–2010 NSFG as were used for the 2002 NSFG, another opportunity to reduce costs was used at the segment selection stage. In many PSUs listed housing units were available for segments selected for the 2002 survey. Using these listed housing units was expected to reduce the cost of listing for the 2006–2010 NSFG. The segment-level selection method described in the following text was modified to allow the maximum overlap between the 2002 and the 2006–2010 NSFG segments within a PSU (13). This overlap reduced the number of segments that required completely new lists of housing units to be prepared.

Within a domain in each PSU, census blocks are ordered geographically by tract, block group, and block number. Segments are selected within a PSU separately for each domain. The selection is systematic with probabilities proportional to the relative size of each segment, the Census 2000 count of occupied housing units in a segment. For each segment in an ordered list of segments in a domain within a PSU, the cumulative number of occupied housing units is calculated for each segment, from the beginning to the end of the list. A selection interval for the systematic selection is computed by dividing the sum of Census 2000 occupied housing units for the domain within the PSU by the number of segments to be selected from the domain.

A random number is generated between 0 and the selection interval. The first selected segment is identified by finding the segment in the ordered list with a cumulated size that exceeds the random number. To obtain the second selection, the sampling interval is added to the random number used to make the first selection, and the next segment in the list with a cumulative size that exceeds this sum is selected. For the third selection, the sampling interval is added to the sum for the second selection, and the next segment in the list with cumulative size exceeding this new sum is selected. The selection continues, adding the sampling interval to the last sum to find the next selection, until the domain allocation of segments is achieved.

For example, suppose that four segments are to be selected in a domain in a particular PSU, and that the Census 2000 count of occupied housing units for all the segments (blocks) in the domain is 100,000. The segments (and blocks) in the domain are sorted geographically by county and tract within the PSU, and a cumulative count of occupied housing units assigned to each segment from first to last. Suppose, for instance, that in the sorted list of segments, the first segment has 75 occupied housing units in the 2000 census, the second has 100, the third 150, the fourth 50, and the fifth 100. Then the cumulative size for each is 75, 75 + 100 = 175, 175 + 150 = 325, 325 + 50 = 375, and 375 + 100 = 475, respectively. The accumulation continues through the sorted list until all segments in the domain have a cumulated count assigned. The last segment in the list has a cumulated count of 100,000.

The sampling interval for the domain is computed as 100,000/4 = 25,000. A random number between 0 and 25,000 is generated, say 425. Then the first selected segment is obtained by finding the first segment with a cumulated count that exceeds 425. In the example, that's segment 5 (with 100 occupied housing units at Census 2000, and cumulated count 475). The interval 25,000 is then added to the random ''start'' of 425: 25,425. The segment with cumulated count that first exceeds 25,425 is then selected as the second segment from the domain. The interval is added to 25,425 again, and the segment with cumulated count exceeding 50,425 is selected. And finally, the interval is added one more time, and the segment within the cumulated count exceeding 75,425 is selected. Four segments have then been selected, each with probability proportionate to the Census 2000 count of occupied housing units in the segment. Only four selections will be made with this procedure, since in the example, for the random start of 425, and for all other random starts, adding the interval to the fourth or last sum produced a number that was greater than the final overall cumulated count of the domain in the PSU, 100,000 in the example.

As noted previously, this selection procedure was modified to maximize the overlap between sample segments in the 2002 and 2006–2010 samples to take advantage of the already available listings in some segments used in 2002. Details of the modification are given in reference 13. This selection procedure chose about 35 percent of the segments in the 2006–2010 NSFG from segments listed in 2002.

The combination of the geographic order and systematic selection produced a sample with a desirable property: the selected segments ''represented'' the entire geographic region of all segments in that domain in the PSU. That is, this process effectively gave the NSFG a geographic stratification of segments within each PSU and domain. And stratification in sample selection, when done in this manner, can reduce sampling variance for estimates obtained from the final sample results. Reduced sampling variance was expected in final sample estimates because of the stratification produced by the systematic selection of segments from the geographically ordered list.

Segments were created and selected the year before the sample PSU was to be used in the national sample. Selected segments were divided into four approximately equal-sized groups, one each for the calendar quarters beginning July 1, October 1, January 1, and April 1.

Field listing of segments

Housing unit lists were prepared for each selected segment. These segment lists came from one of three sources. In about 35 percent of the segments, unused housing unit listings from the 2002 survey were available. These were sent to the field for an ''update.'' In the remaining 65 percent of the segments, a list of addresses for housing units was obtained from a commercial vendor of the U.S. Postal Service's Delivery Sequence File (DSF). Not all segment address lists from the commercial vendor could be efficiently used. In about 15 percent of all the segments (about 20 percent of the segments where addresses were purchased from the vendor), the number of addresses was so low that the addresses were insufficient for subsequent field checking. These were typically in rural areas where either Post Office boxes or rural delivery addresses (without street number and street name) were present. For these segments, an interviewer was sent to do an original ''scratch'' listing (see the following text). Finally, for the remaining 50 percent of all segments, sufficient address counts were obtained from the commercial vendor. The addresses in these segments were sent for an ''update.''

Maps were created to guide an interviewer to the exact location of the blocks in the segment. Maps were generated through commercial software using the Census Bureau's TIGER (**T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing system) files to delineate block boundaries. Typically three levels of maps were created for each segment: a large-scale view showing the location of the segment relative to major highways and streets; an intermediate-scale view showing the segment relative to major streets; and a detailed smaller-scale view showing individual blocks in each segment. Instructions were prepared for either updating or ''scratch'' listing each of the blocks in a segment, including a starting place to begin the listing and a direction to proceed around the block.

The available addresses in a DSF or housing units in segments listed (but not used) for the 2002 NSFG were sorted in a clockwise order, starting in the northwest corner of each block in each segment. Maps and the sorted address list were loaded into a listing application on the interviewers' laptop computer. Interviewers visited the addresses in the calendar quarter before the segment was scheduled to be used in the sample. Interviewers determined whether the housing units or addresses in the sorted list were physically present in the segment. If an address was not present, interviewers deleted it from the list. Interviewers also added housing units to the sorted list when found in the segment but not appearing on the list. Interviewers also reordered the list when they found that the list order did not match the physical order of the addresses in the segment.

For the remaining 20 percent of the segments, where the DSF addresses were too few to justify preparation of an update, a ''scratch'' listing was done. Interviewers were sent to visit the segment with the segment maps and a blank listing provided on their laptop computer. Interviewers visited the segment and listed all housing units within the segment in the blank listing. They followed the same order as used for the update listing, starting in the northwest corner of each block in the segment and moving around the block in a clockwise manner.

The update and ''scratch'' listings were downloaded to ISR as completed. At this point, all listings were considered to be housing units (as opposed to DSF addresses that are not housing units until verified in the field listing). Each segment is checked for completeness and consistency before the third-stage selection of housing units.

Third stage: Selection of households

The third-stage random selection of housing units is made from the segment housing unit list. A listing of an individual housing unit at this point is sometimes referred to as a sample ''line.'' In order to sample lines and assign them to field data collection, a within-domain sampling rate was determined to meet allocation for the interviewer's monthly sample.

Screening and missed housing units

The selection of housing units is continued in the household screening operation in the field. Screening consists of a short questionnaire administered at the doorstep of every housing unit selected for the sample.

Based on ISR experience with listed housing units, 15 percent are expected to be unoccupied or not actually housing units (for example, a housing unit converted to a commercial building). In addition, among occupied housing units, the March 2004 and 2005 Current Population Survey (CPS) indicated that 38 percent were expected to have no eligible persons ages 15–44 years of age.

Screening was the process to determine whether the selected housing units were occupied, and then whether any eligible persons resided in the occupied housing unit. On average, an interviewer had 15 selected housing units from every 100 that were unoccupied, and among the 85 occupied housing units she was expected to find 62 percent (about 53 of the 100 selected housing units) with one or more persons ages 15–44 years.

During the screening phase of the survey, interviewers were trained to check again for housing units that may have been missed in the updating or ''scratch'' listing process. Such missed housing units may occur when an interviewer overlooked a structure with one or more housing units, or missed a part of a structure that was a separate housing unit. Missing units may also occur if a housing unit is constructed since the listing took place.

Interviewers were equipped with a sample management system (SurveyTrak) on their laptop computers that contained all housing units in each segment. The procedure for handling missed housing units in the field was as follows: at each sample housing unit designated in the SurveyTrak listing, interviewers checked to be sure that all housing units following the selected housing unit on the list were present, and checked for mail boxes, doors, or utility meters that might indicate a unit that was not listed. They were instructed to ask screener respondents about any additional housing units in the structure. If an additional housing unit not on the list were discovered between a sample housing unit and the next listed unit, the interviewer added it to the SurveyTrak list, and then attempted a screening interview with the additional units.

When more than two housing units were missing from the SurveyTrak listing, interviewers were instructed to suspend work for that sample housing unit, including contact with the household, and to call the ISR's sampling unit to receive further instructions. Before calling, interviewers were to obtain a list of all additional housing units associated with the sample housing unit. The ISR central office staff then subsample the original and the additional housing units. This creates unequal probabilities of selection of housing units within each domain, and so weighting adjustments to account for the missed housing unit subsampling are incorporated into final weights.

Fourth stage: Selection of eligible persons

The last stage of sample selection was conducted within the household during the screening activities. An adult member of the household was asked to provide a list of all persons living in the household. Information on the gender, age, and race and ethnicity of each person was recorded in the screening portion of the interview (see Figure 4). Interviewers asked additional questions to be sure no one was missed, particularly college students living away from home at a dormitory, fraternity, or sorority. (College students living away from home in their own apartment or housing unit were covered by the household frame, and were not considered to be part of their parents' household.) Dormitory residents were indicated on the household listing.

If no one in the household was between the ages of 15 and 44 years, the household was not eligible to be part of the NSFG sample. If one or more eligible persons were found, the computer-assisted screening system made a selection of one eligible person in the household. That is, an eligible person was selected within each household that had an eligible person.

IntroSC
SC-1 I am (name) from the University of Michigan. Here is my identification card. We are conducting the National Survey of Family Growth for the National Center for Health Statistics, which is part of the U.S. Public Health Service.

This letter explains all about this survey. You may remember receiving this letter and a brochure in the mail recently. Please take some time to read this important information.

Do you have any questions about anything (you have read/I have read to you) about the National Survey of Family Growth?

Enter (1) to continue
Underlying range: 1 to 1, NODK, NORF

	Name	Sex	Age	Race	Hisp	DormRes
HL(1)						
HHL(2)						
HHL(3)						
HHL(4)						
HHL(5)						
HHL(6)						
HHL(7)						
HHL(8)						
HHL(9)						
HHL(10)						
HHL(11)						
HHL(12)						
HHL(13)						
HHL(14)						
HHL(15)						

Figure 4. Within-household screener and household roster

Age and sex	Black	Hispanic	White or other
Female:			
15–19	1.00	1.00	0.84
20–24	0.94	0.94	0.84
25–44	0.87	0.84	0.28
Male:			
15–19	1.00	1.00	0.84
20–24	0.90	0.57	0.47
25–44	0.68	0.47	0.12

Illustration:

Race and ethnicity	Sex	Age	Measure of size	Cumulative measure of size	Random number (from 0 to 3.45)
Black	Female	6	0.00	- - -	
Black	Female	15	1.00	1.00	← 0.95
Black	Female	40	0.87	1.87	
Black	Female	10	0.00	- - -	
Black	Female	21	0.90	2.77	
Black	Female	42	0.68	3.45	

- - -Category not applicable.

Figure 5. Within-household measures of size and illustration of within-household selection

Subsampling by age, gender, and race and ethnicity

The sampling rates used within households for eligible persons varied by age, gender, and race and ethnicity, since the selection is carried out by a computer application. This system allowed SRC and NCHS staff to achieve target sample sizes more precisely in the face of uncertainty about rates of eligible persons in the population.

The within-household selection procedure assigned a ‘‘measure of size’’ to each age-eligible person in the household. The measure of size was 1 of 12 values, 1 for each of 12 age, sex, and race and ethnicity subgroups of interest (see Figure 5 for an illustration). The measures were numbers between zero and one. Larger numbers assigned to a subgroup increased the chances that persons in that subgroup would be selected for interviewing. Thus, larger numbers were assigned to teenagers 15–19 years of age so that larger numbers of teenagers would be selected for interviewing, and sample sizes increased. Slightly larger numbers are also assigned to females to increase the number of females relative to males in the final sample.

The specific measures of size values were developed through a simulated sample selection from concatenated CPS data for 2004–2005 to generate the desired sample sizes by subgroups by age and gender. For each member of a household in the concatenated CPS data file, measures of size were assigned and a subject selected. The measures were adjusted, subgroup by subgroup, to achieve target sample sizes. The rates were then applied to the continuous NSFG within-household selection.

Once each eligible person was assigned a measure of size, the sizes were cumulated. A random number from zero to the sum of the measures in the household was generated by the computer-assisted sample screener application. The first listed person whose cumulative measure of size within the household exceeds the random number is selected. The chance of selection of the person is thus proportionate to their measure of size in the household.

The lower panel of Figure 5 illustrates how the selection would be executed in a hypothetical household with six persons: the black female age 15 has a measure of size of 1.0, which exceeds the random number of 0.95, so she was chosen to be interviewed. By assigning larger measures to teens (i.e., those ages 15–19 years) and females within the household, larger samples of these individuals are achieved. The selection probabilities within the household depend on the distribution of measures of size for other eligible persons in the household.

Implementing the Sample in a Responsive Design Context

Surveys with high response rate goals and limited budgets such as the NSFG need a way to stay informed on the level of key survey design parameters such as completed interviews, eligibility rates, response rates, expenditures, and interviewer productivity. In most surveys, the information systems that provide such data are designed to provide some data daily and other data at the end of the data collection period. There is seldom an opportunity to make changes to a survey design based on this information between the start and end of data collection.

In contrast, the NSFG's information systems allowed survey design changes throughout the data collection. The information systems provided daily data on such features as how many interviewer hours and how much money were being spent on data collection, what areas and interviewers were having good and poor results, and what types of nonresponse were most prevalent in an area. ISR uses a system called SurveyTrak to provide this daily information.

NSFG project staff used SurveyTrak and other information systems to manage data collection to keep within budget and meet survey data quality targets. The key variables in implementing responsive design are the information systems that supply the data on a daily basis and the trained staff to use them. Thus, any survey done at any pace could implement responsive design. A continuous interviewing system, which spreads the data collection tasks over time, however, offers more opportunities to use information systems to adapt sample size and data collection procedures to survey conditions.

The 2006–2010 NSFG design relied on a set of five sampling levels: the PSU, the segment, the housing unit, the person within the housing unit, and the second-phase (or ‘‘double’’) sample. The sample levels permitted design changes at regular intervals throughout data collection, where sample sizes could be varied depending on survey information about eligibility rates, response rates, and interviewer performance. These features constituted most of the design elements of the survey's ‘‘responsive design’’ that were used to control sample size and response rates.

The first step of the responsive design process was to set sample size at the **first level (the PSU)** on the basis of an annual review of SurveyTrak data. Each annual quarter sample was fielded in one data collection year starting on about July 1 of 4 successive years. These national samples could have sample size targets computed based on the most recent data available on expected interviewer workload and past interviewer performance in the same PSU, or a similar PSU, PSU-specific eligibility rates, and past or estimated PSU-level response rates.

Within an annual national sample, there was a further selection at the **second level** within the selected PSUs: the segment selection. As described previously in more detail, census blocks within the PSU from the 2000 Census of Population and Housing are divided into four groups, or domains, on the basis of the concentration of black and Hispanic populations within each block. Separate samples were selected from each domain to allow 12 or more segments to be selected in most PSUs. (The eight largest PSUs often had samples greater than 12 segments, and some smaller nonmetropolitan PSUs had fewer.) These 12 segments were selected according to probability sampling procedures. The 12 segments were randomly allocated into four sets of approximately three segments each, and each set was assigned to a sample for one 12-week quarter, released starting July 1, October 1, January 1, or April 1 in each data collection year. As for the annual national quarter samples, the size of sample from each segment in each quarter could be adjusted based on SurveyTrak data to reflect interviewer workload and performance, expected eligibility rates, and expected response rates.

The **third level** sample selection was a *housing unit sample* within the segments selected for use in a given calendar quarter. In each selected segment, one of two procedures was used to list housing units before selection. In the ‘‘update’’ segments, addresses were obtained from a commercial source or an unused list prepared for the 2002 NSFG, visited by interviewers to check and correct list accuracy, and selected by central office staff at ISR for screening and interviewing. In the ‘‘scratch’’ segments, interviewers visited the segment and listed all housing units, and a sample of the updated or listed addresses was selected from the interviewer list by ISR central office staff. The sample selection rates and cluster sizes could be varied across segments depending on the housing unit yield of the listing operation in order to yield a number of housing units in each interviewer's assignment for the calendar quarter to match interviewer efficiency (hours per interview, described in the following text) as well as expected response, occupancy, and eligibility rates. The determination of selection rates within PSUs, and thereby within segments, is described in detail in the next section.

The **fourth level** of selection was the random choice of one person 15–44 years of age within the household. Interviewers visited selected housing units in assigned segments starting at the beginning of the calendar quarter. A household roster was generated containing a list of all persons who usually resided in the household. If one or more of these was 15–44 years of age, one age-eligible person was selected with varying probabilities in each of these households in the *person level sample*. If the selected person was present at the time of this screening visit, the interviewer attempted to interview him or her. Otherwise, interviewers collected information to set a time to revisit the household to interview the selected person. Sample persons were offered $40 as a token of appreciation at the time of selection. The within-household selection probabilities could be varied from one calendar quarter sample to the next to achieve target sample sizes for key subgroups defined by age and gender.

A **fifth level** of selection that provided an opportunity to respond to design circumstances as well occurred at the 10th week in each calendar quarter. At that time, some selected housing units had not been successfully screened or had been screened but a completed main interview has not been obtained. A sample, typically about one-third, of these remaining addresses were selected for *Phase 2 sample* interviewing during the last 2 weeks of the calendar quarter. Effectively, interviewer assignments were reduced so that interviewers concentrated on a smaller number of housing units and selected persons for the final 2 weeks of data collection in the 12-week quarter. This within-quarter level of selection was particularly useful to control final response rates and costs for the overall sample in the calendar quarter.

SurveyTrak data were again used to adjust the sample size in each segment. Prediction models, based on SurveyTrak data generated daily, were used to group housing units that had not been completed by the end of week 10 into sets with varying levels of predicted propensity to respond. Housing units were selected from each group, and the selected housing units constituted the Phase 2 sample released to each interviewer at the beginning of the 11th week of data collection. Somewhat larger tokens of appreciation were used in Phase 2 to encourage higher levels of participation. At the end of the quarter, a weighted response rate was computed across the two phases of data collection and data generated to allow assessment of the extent to which nonresponse bias was being reduced by the addition of Phase 2 interviews. A more detailed description of the Phase 2 design and procedures can be found in the following text, in the section, ''Sampling in Two Phases.''

Sample Release Design

With each annual and each calendar quarter sample released, key design parameter estimates were altered to change sample allocation across segments for each calendar quarter. The sample sizes were varied by PSU and by quarter to respond to changing survey conditions. Thus, the sample size of lines or persons in each quarter was not fixed—it could be adjusted based on the most recent available data.

The primary parameter in the allocation at the PSU and segment level was the interviewer workload. Interviewers were recruited and hired to work an average of 30 hours per week, or approximately 360 hours in a 12-week quarter. This management model required an allocation of sample across PSUs and segments that was not standard. In many standard survey allocations, the number of sample lines to be chosen from a segment and a PSU is determined on average across all PSUs, or PSUs of a similar type (such as NSR). In the 2006–2010 NSFG, this standard practice would have allocated approximately the same number of housing units to be selected across segments within the same domain. Interviewers worked to complete the sample lines, regardless of expected response and eligibility rates and their expected efficiency.

In the 2006–2010 NSFG, however, a more efficient management model was used for interviewer employment. Interviewers were recruited to work an average of 30 hours per week throughout 48 weeks of the year. (Interviewers in nonrotating ''super 8'' PSUs are promised 30 hours per week for as long as their performance met ISR standards, since the ''super 8'' PSUs were in the sample for all 4 years of the design.) This hiring, recruiting, and employment model forced a nonstandard method of allocating lines across segments. The management target is a fixed number of <u>hours of work</u> in a calendar quarter, instead of a fixed number sample housing units to be visited.

Within an expected 360 hours in a 12-week period, interviewers update or prepare ''scratch'' listings for the segments allocated in the next calendar quarter, screen selected lines, and conduct main interviews. Interviewers had in their work assignments varying survey conditions that could make them more or less efficient within the 360 hours. The conditions varied by the nature of the communities in which they work, which in turn affects parameters such as the number of hours required to complete an interview (i.e., the hours per interview for PSU α at calendar quarter t, $HPI_{\alpha t}$); the housing unit occupancy rate ($\hat{O}_{\alpha t}$); the proportion of occupied housing units with one or more persons ages 15–44 (the eligibility rate, $\hat{E}_{\alpha t}$); and the combined screener and main interview response rate ($\hat{R}_{\alpha t}$).

Prior to each calendar quarter, study staff asked interviewers to provide an estimate of the number of hours to be worked in the next quarter. The sample line assignment process then allocated hours equally over the segments that would be assigned to the interviewer in the quarter, say $H_{\alpha t}$ for the α^{th} PSU at calendar quarter t. Estimates $HPI_{\alpha t}$, $\hat{O}_{\alpha t}$, $\hat{E}_{\alpha t}$, and $\hat{R}_{\alpha t}$ are prepared for each PSU as well. The estimation took into account variation in the achieved values of these parameters by region (that is, Northeast, South, Midwest, and West) and by segment domain, generating 16 values of these parameters. For each PSU in the sample, the number of housing units or lines to be allocated for the forthcoming calendar quarter t was computed as

$$L_{\alpha t} = \frac{(H_{\alpha t} / HPI_{\alpha t})}{(\hat{E}_{\alpha t} \times \hat{O}_{\alpha t} \times \hat{R}_{\alpha t})}.$$

These housing unit or line estimates $L_{\alpha t}$ were adjusted after review by study staff to account for interviewer or PSU conditions that departed from expectations for region and domain.

This allocation led to variation in probabilities of selection of housing units across segments within and among PSUs. The variation was compensated for in the weighting process (described in the following text), although the added variability in sample weights from varying line probabilities at the segment level has the potential to increase the variability of survey estimates.

Thus, for the sake of management efficiency, the NSFG sample was designed around interviewer productivity. The calendar quarter allocation of sample lines per segment allows study staff to adjust sample size to account for variation in eligibility, occupancy, and response rates, as well as hours per interview, each calendar quarter.

Sampling in Two Phases

Each NSFG calendar quarter consisted of two phases. In the first 10 weeks of the quarter, interviewers had to screen selected lines in assigned segments, conduct main interviews in households with eligible persons, and update or prepare ''scratch'' listings for the segments allocated in the next calendar quarter. After 10 weeks of data collection, sample lines (or cases) remained that had not been successfully screened and sample persons who had not yet completed the interview. If the data collection was halted at the end of 10 weeks, these unscreened households and noninterviewed persons would contribute to nonresponse bias. A ''double or two-phase sample design'' (14) was instituted for the remaining 2 weeks of the quarter as a device to reduce the nonresponse bias in survey statistics. The elements of a ''two-phase design'' (as Hansen and Hurwitz in (14) coined the term) are:

A. The design and implementation of a survey design on a large sample (labeled the ''first phase'').
B. The selection of a probability sample of the nonrespondents from the first-phase sample.
C. The use of a different participation protocol for the second phase.

There are two impacts of a two-phase design. First, if the second-phase protocol is successful in measuring (i.e., interviewing) 100 percent of the *sampled* nonrespondents from the first phase, nonresponse bias is eliminated. In practice, no subsample of nonrespondents attains a 100 percent response rate and thus some non-response bias remains, but the bias is expected to be reduced by the capture of data from the first phase nonrespondents. Second, the cases sampled into the second phase that are successfully interviewed are assigned new selection weights (reflecting the fact that they must ''represent'' the nonselected nonrespondents). This additional weight component generally increases the variance of the estimates.

Two-phase designs are increasingly attractive to survey researchers because they offer a way to control the costs at the end of a data collection period, while addressing concerns about nonresponse rates and errors. In face-to-face surveys, at the end of the data collection period, large costs are incurred for travel to sample segments to visit only one or two sample units, usually those extremely difficult to contact in prior visits or repeatedly displaying some reluctance to grant the survey request. By restricting these expensive visits to a sample of the nonrespondents at the end of the study, a more cost-effective method limits costs while addressing the need to increase response rates.

In the 2006–2010 NSFG design, a subsample of nonrespondents was chosen for weeks 11 and 12 based on study of the history of the first 10 weeks' sample. Study staff developed response propensity models to predict the probability that a given case yields a completed screening interview or a completed main interview (see reference 11 for details of the propensity models). Within a PSU, one of the three segments was deleted at random. The active nonresponse cases in the two remaining segments were grouped by type (screener or main interview) and four categories based on their estimated propensity as of the conclusion of the first 10 weeks. A disproportionately allocated sample of nonresponse cases was selected across these four strata, with higher probabilities of selection from strata with higher propensities. These selected lines and persons were then released to interviewers for Phase 2 data collection in the last 2 weeks of the calendar quarter.

Under a responsive design theoretical perspective (15) that guided the NSFG's design and fieldwork, study staff sought a Phase 2 interview recruitment protocol that was distinctive from that used in Phase 1. Such a distinction is necessary (but not *a priori* sufficient) to attract sample persons who did not find the Phase 1 protocol effective, and thus to increase response rates and reduce bias in the sample data. With the approval of two Institutional Review Boards and the Office of Management and Budget, the Phase 2 recruitment protocol involved the following components:

A. Limited use of proxy respondents (such as neighbors) for the screening interview was allowed in weeks 11–12 of each quarter.
B. A prepaid $5 token of appreciation payment (versus none) for cases that had not yet completed the screening interview.
C. A prepaid $40 token of appreciation for the main interview (versus a $40 token provided after the informed consent was signed).
D. A promised additional $40 token of appreciation for a completed main interview.

Sample Weighting

Overview of Sample Weighting

A simple random sample in which response rates and coverage are the same in every subgroup would create a ''scale model'' of the population. However, many survey samples are not ''scale models'' in that sense. If a ''scale model'' of the population is selected, smaller groups in the population would have too few observations in the sample to provide adequate precision for characteristics of interest for those groups. As a result, survey samples often deliberately select groups at higher and lower rates to over represent smaller groups in the sample. This allows analysts the opportunity to answer key survey questions for the total population and for those small but often important groups of the population.

For example, in the NSFG, non-Hispanic black women and men were sampled at a rate that made them about 20 percent of all respondents in the sample, even though they are approximately 13 percent of the population 15–44 years of age. Hispanic women and men and teenagers of all races were also sampled at higher rates in the 2006–2010 NSFG yielding about 20 percent of the sample interviewed from each group. ''Sampling weights'' adjust for these different rates.

A respondent's sampling weight can be interpreted as the number of persons in the population that he or she represents. For example, if a woman's sampling weight is 8,000, then she represents 8,000 women in the population. For the NSFG, the fully adjusted sampling weights are assigned to each respondent and consist of three factors. The first factor is the inverse of the probability that the case was selected. For example, in NSFG the first stage of the weight construction is the probability of selection for each sample person. If the probability is 1 in 6,000, then the initial sampling weight is 6,000. The second factor is an adjustment for nonresponse, calculated separately and based on the probability of completing a screener and the probability that a completed screener results in a completed interview. If that response rate is 75 percent, or 0.75, then 6000/.75 = 8,000, so the nonresponse adjusted weight would be 8,000.

The third factor is termed ''poststratification'' and is an adjustment to control totals of the number of persons by age, sex, race and Hispanic origin, provided by the U.S. Census Bureau (and, for military personnel living in a household outside of a military facility, the Defense Manpower Data Center). If, for example, it is necessary to increase the weights for cases in our example cell by 5 percent to agree with the poststratification totals, then the poststratified weight would be (8,000) × (1.05) = 8,400.

More thorough discussion of weighting procedures appears in the ''Detailed Weighting Procedures.'' The use of weights in estimation requires that standard error estimates reflect these weights; see ''Overview of Variance Estimation.''

Detailed Weighting Procedures

The 2006–2010 NSFG was designed to estimate the number of women and men with particular characteristics in the U.S. household population. For example, the NSFG can provide estimates of the number of women who have ever had sexual intercourse for the total population and for groups such as teenagers 15–19 years of age, men 20–24 years of age, or Hispanic women 25–44 years of age.

The 2006–2010 NSFG deliberately oversampled specified subgroups of the population. It was also subject to nonresponse and noncoverage error. Sampling weights were assigned to every person in the public-use data files. These weights must be used to compensate for unequal probabilities of selection and for under-representation of subgroups due to nonresponse or noncoverage. That is, the sampling weights are designed to produce estimates from the sample that correct for oversampling, nonresponse, and noncoverage.

The weights were constructed in four steps. In the first step, a set of weights is constructed to account for deliberate or designed over- and under-sampling features of the NSFG sample. These weights are combined into a single ''base weight'' to adjust for unequal probabilities of selection across all such sampling features. This is sometimes called a ''probability-of-selection weight.''

In the second step, an extensive set of nonresponse models was developed to predict the probability of response to the screener for all households that contained eligible persons and for all selected eligible persons within successfully screened households. These models used predictors obtained from the 2000 Census of Population and Housing, from the housing unit listing process, from interviewer observations collected at each contact with the household, from household call records, from household composition, and from several additional household and selected person characteristics. The inverse of the predicted probabilities from the screener and main interview nonresponse models were used as adjustment factors applied to the base weights.

Third, nonresponse adjusted base weights were summed within each of 36 groups—six age, two gender, and three race and ethnicity categories. The cumulated weights were compared with counts that combine independent Census Bureau population projection estimates for each of the 36 groups and U.S. Department of Defense counts of military personnel in each of the 36 groups living off-base. Poststratification weights are computed for each cell as the ratio of Census plus Defense Department counts to the cumulated sample weights in each subgroup. The poststratification factors are applied to the individual nonresponse adjusted base weights.

Fourth, the poststratified weights were inspected to detect extremely large or small weight values. A small number of large or small weight values are ''trimmed'' to reduce the overall variability of the weights. The trimming procedure is designed to minimize the impact on weighted estimates, while at the same time reducing the variability of the weights as much as possible. The poststratification factors for the respective cells of each trimmed weight were recomputed after trimming. The weights will be released with the data files.

Inverse probability selection weighting

The NSFG sample management system preloaded components of the probability of selection of each sample line onto the SurveyTrak record. Each sample person was selected through PSU, segment, housing unit, and person within selected household selection procedures. In addition, the Phase 2 sample selects a subset of the remaining nonrespondent cases for intensive follow-up for the last 2 weeks of fieldwork.

Primary sampling units

Each of the 110 PSUs was selected with probabilities proportionate to the number of occupied housing units. The probabilities were computed at the time of selection and stored on the SurveyTrak record for each sample line.

Segment selection

Blocks were selected within each domain within PSUs with probabilities proportionate to the Year 2000 census count of the number of occupied housing units in the block. In all PSUs, segments in domains 2, 3, and 4 are chosen with higher probabilities of selection than for those in domain 1. Probabilities of selection for segments within domains are proportionate to the estimated number of households in the segment. The probability of selection of each block is computed at the time of selection and stored in the SurveyTrak record for each line for later use.

Housing unit selection

Not all housing units in sample segments were selected for the NSFG sample. Segment housing units were subsampled in Phase 1 to achieve a target number of sample housing units or lines. The subsampling probabilities of selection for housing units vary by segment and are recorded on the SurveyTrak record.

Person selection

Once sample housing units or lines were selected and released through the ISR sample management system to interviewers in each PSU, interviewers visit sample housing units to determine if any eligible persons reside there. The interviewer completed a household roster in the Blaise® instrument, recording age, gender, and race and ethnicity for each member of the household. If one or more persons 15–44 years of age lived there, a random selection was made of one eligible person per household—with the chances of selection varied to increase the selection of teens (ages 15–19), women, and blacks and Hispanic persons. The household roster and chances of selection were recorded in the Blaise® household record.

Phase 2 sample design

A final component of sample selection was introduced in Phase 2 sampling in the final 2 weeks of the calendar quarter data collection. On average, a sample of two or three segments in a calendar quarter was chosen in each PSU. Nonresponding housing units in the selected segments (those which have not reached a final disposition after 10 weeks of data collection) were divided into strata on the basis of type (screener or main) and predicted probabilities of obtaining a completed interview. A Phase 2 selection of segments and of nonresponding housing units was chosen, with higher chances of selection assigned to those nonresponding housing units with higher estimated response propensities. The varying chance of selection for segments and housing units in the second phase selection were retained for subsequent weighting.

Probability of selection and weight

The probability of selection of each sample person can thus be computed using the probabilities of selection for PSUs, segments, sample line, within-household selection, and Phase 2 subsampling of replicates.

Let $M_{h\alpha}$ denote the size measure of the α^{th} PSU in stratum h, the number of occupied households in the PSU in 1990. Let $M_{h\alpha\beta}$ denote the size for the β^{th} segment in the $(h\alpha)^{th}$ PSU, where the size measure for each segment is the number of occupied housing units in the 2000 Census. Also, let $L_{dh\alpha}$ denote the desired number of sample lines for the d^{th} domain in the $(h\alpha)^{th}$ PSU. Finally, let $\pi_{2,h\alpha\beta}$ denote the combined Phase 2 within-household selection probability for the $(h\alpha\beta\gamma)^{th}$ sample person.

The probability of selection of the $(h\alpha\beta\gamma)^{th}$ eligible person (where γ denotes the γ^{th} sample person in the $(h\alpha\beta)^{th}$ segment) is computed as

$$\pi_{h\alpha\beta\gamma} = \left(\frac{M_{h\alpha}}{\sum_{\alpha=1}^{a_h} M_{h\alpha}}\right) \bullet \left(\frac{M_{h\alpha\beta}}{\sum_{\beta=1}^{b_{h\alpha}} M_{h\alpha\beta}}\right) \bullet \left(\frac{L_{dh\alpha}}{M_{h\alpha\beta}}\right) \bullet \pi_{2,h\alpha\beta}$$

The base weight compensating for unequal chances of selection for the $(h\alpha\beta\gamma)^{th}$ eligible person is the inverse of this probability of selection,

$$w_{h\alpha\beta\gamma} = w_{1i} = \pi^{-1}_{h\alpha\beta\gamma}.$$

This variation in base weights due to the over- and under-sampling within households has the potential to increase the variance of estimates. A summary factor of the potential increase in variance, computed under the assumption that the weights are uncorrelated with the values of the characteristics being estimated, is estimated for each of the 12 age-gender-race and ethnicity subgroups of greatest interest. The factor 1+L is a measure of the relative variance of the weights themselves, and is computed as

$$1{+}L = n\,\frac{\left(\sum_{i=1}^{n} w_{1i}^2\right)}{\left(\sum_{i=1}^{n} w_{1i}\right)^2}.$$

This factor is an approximate potential relative increase in the variance of estimated means that can be attributed to the distribution of the weights. The full effects of these increases are reduced later in the weighting process through trimming of the largest final weight values.

Nonresponse adjustment

Nonresponse in the 2006–2010 NSFG occurred at both screening to identify sample eligible persons in sample households and at the main interview among selected eligible persons. If there was nonresponse at the initial contact for the screener interview, there was little or no information about the address. The main interview nonresponse occurred any time after the conclusion of screening—that is, after a sample person had been selected. The main interview nonresponse therefore had information about household composition and race and ethnicity of the selected person, as well as a few additional variables that were obtained during household screening.

Nonresponse adjustment for the NSFG was implemented under an assumption widely used in the adjustment of survey data, missing at random (MAR). That is, within subgroups of sample units (housing units in the screener, selected eligible persons in the main interview), it is assumed that the nonrespondents are a random sample from all the units in the sample. A nonresponse weighting adjustment developed under this assumption is computed as the inverse of an estimated response rate within a subgroup. This sample-based weight represents an adjustment that, under the MAR assumption, substitutes for a probability of selection in the response process. Thus, as for unequal probability weighting, the inverse of the predicted probability of response serves as an adjustment factor.

There are many methods for estimating response rates (16,17). Most are alternative ways of estimating probabilities of responding under MAR. The weighting class method divides the sample into weighting classes across which response rates are expected to vary, and across which the characteristics of sample persons are expected to vary. An alternative method is to estimate probabilities of response through a propensity model, such as in a logistic regression model. The 2006–2010 NSFG uses this latter logistic regression method, computing separate models for screening and main interview nonresponse.

Let S_i denote a zero-one indicator for whether a sample address is successfully screened to determine whether eligible persons lived in the household, where 1 denotes successful screening. (S_i is not defined for sample addresses that were not occupied.) The screening logistic regression model for all occupied sample addresses is

$$\lambda_{si} = Pr(S_i{=}1|X_i) = (1+exp(-X_i'\beta))^{-1}$$

where X_i is a vector of predictor values for the i^{th} occupied housing unit and β is a vector of logistic regression coefficients. The coefficient values are estimated through standard maximum likelihood methods, and predict the logit

$$\hat{\lambda}_{si} = (1+exp\,(X_i'\hat{\beta}))^{-1}.$$

This logit in turn is used to obtain a predicted probability of successful screening as

$$\hat{\pi}_{si} = \frac{exp(-\hat{\lambda}_{si})}{(1+exp(-\hat{\lambda}_{si}))}.$$

Similarly, let R_i denote a zero-one indicator for main interview response for the i^{th} successfully screened occupied housing unit, equal to 1 when the selected eligible person has a completed interview, and 0 for selected eligible persons who did not complete an interview. (R_i is not defined for sample addresses that are not occupied or are not successfully screened.) The main interview logistic regression model for selected eligible persons is

$$\lambda_{mi} = Pr(R_i{=}1 \mid S_i{=}1, Z_i) = (1+exp(-Z_i'\gamma))^{-1}$$

where Z_i is a vector of predictor values for the i^{th} selected eligible person and γ is a vector of logistic regression coefficients. Standard maximum likelihood methods yield

$$\hat{\lambda}_{mi} = (1+exp\,(Z_i'\hat{\gamma}))^{-1}.$$

The predicted probability of successful screening is

$$\hat{\pi}_{mi} = \frac{exp\,(-\hat{\lambda}_{mi})}{(1+exp\,(-\hat{\lambda}_{mi}))}.$$

A model is fit to existing data separately for screening and for main interviews among successfully screened addresses. The overall response propensity is then estimated for each housing unit based on the product of two predicted probabilities, $\hat{\pi}_{ri} = \hat{\pi}_{si} \times \hat{\pi}_{mi}$. This predicted probability is computed for all selected eligible persons.

The nonresponse weight is the inverse of the predicted probability $\hat{\pi}_{ri}$, but only for completed interview cases. In particular, $w_{ri} = \hat{\pi}_{ri}^{-1}$ if $R_i = 1$ and $S_i = 1$, and $w_{ri} = 0$ otherwise.

Unlike the 1995 NSFG, which had the National Health Interview Survey data set for each sample person to use as potential predictors in response propensity models, the NSFG now has a more limited set of geographic and operational variables to use as predictors for the screening and main interview models. Preliminary modeling of screener and conditional main interview response propensity was performed in the estimation of segment level expected completed interview counts for the Phase 2 sample selection (see reference 11). The preliminary models employed a number of predictors:

1. Counts and rates for the segment from which the housing unit is selected, derived from 2000 census data for the blocks in the segment.
2. Data obtained from observations made at two levels for each housing unit: characteristics of the segment and housing unit recorded by the interviewer listing the segment.
3. Respondent behavior recorded by the interviewer at each contact with anyone within the housing unit.
4. Operational measures, such as number of calls to a housing unit, number of calls to the sample person, and interviewer response rate.
5. For the main interview propensity model, data drawn from the household roster and other data collected in the screening interview.

These sets of variables were used as predictors in the response propensity models for screener and main interview. The census variables, segment and housing unit observations, screener contact observations, and operational variables are being examined in stepwise logistic regressions for the screening indicator variable S_i. The same predictor variables plus main interview contact observations, household composition, and other variables available from the screening interview are being examined in stepwise logistic regressions for the main interview response indicator R_i.

Screener response propensity model

The predictors in the screener propensity model are a set of variables theoretically important for response propensity. The set of variables to be used in the screener propensity model is shown in Table B.

Substantial variation in predicted probabilities generates considerable variation in response propensity weights. A common practice in survey estimation is to reduce this variation by grouping predicted values into classes, and then using some middle value to represent the entire group's predicted values. This method effectively imposes a constraint on the underlying propensity model. The predicted probabilities are grouped by deciles of the predicted probabilities. All completed screener cases in the lowest predicted probability decile are assigned the value of the median of the predicted probabilities in the decile. The remaining 90 percent of the values are not changed. This response propensity revision reduced the variation to a range from 0.39 to 0.98. The main interview nonresponse weight W_{mi} is computed as the inverse of the predicted probability, or the inverse of the median of the lowest decile.

Main interview response propensity model

The main interview response propensity model development has more predictor variables than the screener propensity model because a larger number of predictors is available, and a broader theoretical framework can be used for weight development. The stepwise logistic regression does not select such variables as age, gender, race, and ethnicity as predictors of response propensity for the main interview model. However, all are known to be predictors of the values of the fertility and other characteristics observed in the interview.

Theoretically, there are two general approaches to developing weights to adjust for nonresponse. The response propensity approach seeks to find a set of subjects whose predicted probabilities of response are identical, or nearly identical, and then assign to the respondents in the set a common weight that compensates for the nonrespondents in the set. Because there is no variation among the cases in the set on the propensity to respond, there is no association between characteristics of interest and response propensity. That is, nonresponse is independent of the characteristics of interest. At the same time, individuals with different propensities (that is, in different sets) who also have different values of the characteristic will generate association between propensity and substantive measures. The inverse propensity weights thus allow the association between propensity and substantive values to be used to adjust the estimates.

A second approach is the predictive approach. If a group of sample persons with the same or similar propensities has the same value of a substantive measure, nonresponse is independent of the substantive measure. Models that predict substantive measures can be used to group respondents together based on predicted values of substantive variables. Inverse propensity weights within these groups will then allow the association between propensity and substantive measures to be used to adjust estimates.

The predictive approach is difficult to implement in practice because good predictors of substantive measures are not usually available for nonresponse adjustment for both responding and nonresponding cases. However, in the NSFG, age, sex, race, and ethnicity are known for all selected sample persons and are important predictors of substantive measures. Thus, while the response propensity model stepwise regression screening may not select these predictive variables, it is important to include them in some form in the propensity models. That is, a combination of response propensity and predictive models were used for the main interview nonresponse adjustment.

Table B. Screener propensity model predictors, 2006–2010 National Survey of Family Growth

Predictor name	Predictor description
LRESIDENTIAL	All housing units in sample segment are residential (yes/no)
MANYUNITS	Address in a structure with multiple housing units (yes/no)
EVERQUEST	Informants at address asked one or more questions during one or more contacts (yes/no)
EVERRESIST	Informants at address made statements indicating reluctance to be interviewed (yes/no)
NCALLINFOLD	Total number of calls made to address
NCONTACT	Total number of contacts with household
EVENCALLPER	Percentage of calls made during evening hours
EVERSTATE	Informants at address made statements delaying interview (yes/no)
CHILD_LT15	Interviewer assessment of whether anyone in the household is under 15 years of age (yes/no)
IWER_EXP	Number of days since interviewer was hired by UM-ISR
BILINGUAL	Interviewer speaks English and Spanish (yes/no)
REGION	Geographic region (Northeast, Midwest, South, West)
PSU_TYPE	Type of PSU (Super eight, other large metropolitan, and smaller metropolitan or nonmetropolitan)
DOMAIN	Segment has 1) more than 10% population Black, 2) more than 10% Hispanic, 3) more than 10% Black and Hispanic, or 4) all other
QUARTER	Calendar quarter of data collection (1 to 16, where quarter 1 is approximately July 1 to September 30, 2006, quarter 2 October 1 to December 31, 2006, etc.)
SCRNPROB	Interviewer assessment of probability (high, medium, low) of obtaining screener interview made at week 7 of each quarter
SCRN_PHASE2	Address selected for Phase 2 sample as a screener case (yes/no)

In addition, interviewers were required to provide subjective assessments of several household or personal characteristics at the first visit to the sample housing unit. One of these, an interviewer's assessment of whether there is an active sexual relationship in the household (i.e., a married or cohabiting couple, as indicated by a screener informant referring to ''my husband'' or ''my partner''), has proven to be correlated with key variables and thus useful in a predictive model.

A number of different models fit to the data will employ a reduced set of propensity predictors in subclasses defined by age and interviewer-assessed active sexual relationship (Table C). Within these groups, separate models with similar predictors are to be used in each subgroup propensity model. The level of change in the logarithm of the likelihood from the ''null model'' (containing only an intercept as the predictor to the model) will be observed to determine if these models explain substantial amounts of variability in propensity. Thus, the main interview propensity estimates will most likely be based on subgroup models selected through stepwise regression.

Combined nonresponse weight

As a final step in the construction of the nonresponse adjustment weight, the screener nonresponse weight w_{si} is multiplied by the main interview nonresponse adjustment w_{mi} to obtain the final nonresponse adjusted weight, $w_{ri} = w_{si} \times w_{mi}$. No trimming of this final weight value is expected to be needed.

Poststratification

The final stage in the weighting process is the adjustment of weighted sample values to outside distributions. Preliminary study indicated that there are few external data to which the NSFG data could be benchmarked in a calibration approach.

A calibration to external population estimates by age, gender, and race and ethnicity adjusts the nonresponse adjusted weight w_{2i} to an estimate of the civilian noninstitutionalized population in the group provided by the Census Bureau plus counts for military personnel living off base obtained from the Defense Manpower Data Center. Across 36 age, gender, and race and ethnicity cells, a poststratification weight w_{3i} is computed as the ratio of the combined Census and Defense counts to the sum of the nonresponse adjusted weight w_{2i} in each cell. A preliminary final weight $w_{4i} = w_{3i} \times w_{2i}$ is being computed for each of the completed interview cases.

Weight trimming

The distribution of the preliminary final weight is being examined in each of the primary 12 subclasses as well as by groups formed by age, gender, race and ethnicity, and two-way combinations of these variables. Potential increases in variance (the factor 1+L presented previously) greater than 2.0 may be observed in some of the 12 subclasses, suggesting that individual weight values may be contributing to a substantial variation in the weights in the subclass.

Considerable reduction of the variability in the weights can be achieved by a reduction of a few extremely large weight values. These kinds of *ad hoc* weight trimming procedures are used on occasion in many surveys under the assumption that reduction in weight variation will reduce standard errors of survey estimates while at the same time leading to inconsequential changes in the survey estimates themselves. The trimming process first trimmed the second-phase probability weight to be no more than 2.0. Then individual weight values and their components (that is, w_{1i}, w_{2i}, w_{3i}, w_{4i}) for each of the 12 subclasses were listed for those cases where the preliminary final weight value is more than two times the mean weight value for the subclass. The nonresponse adjusted base weight for cases with the largest weight is trimmed to the next largest nonresponse adjusted base weight value, and poststratification weights recomputed following this trimming. The potential increase in variance factor 1+L is recomputed, and weighted estimates for as many as 20 key variables (variables monitored during data collection to assess changes in value as data are collected) are also computed using the trimmed and untrimmed poststratified weights. The relative change in the 1+L value is compared with the relative change in the weighted estimates for the cell where trimming occurred. If the change 1+L from untrimmed to trimmed weights is large, and the relative change in the trimmed and untrimmed weighted estimates is small, the trimming is accepted as satisfactory. Trimming stops in subclasses when the relative change in weighted estimates becomes large from one trimming step to the next, and the change in the factor 1+L is small.

Table C. Main interview propensity model predictors, 2006–2010 National Survey of Family Growth

Predictor name	Predictor description
LRESIDENTIAL	All housing units in sample segment are residential (yes/no)
LSAFECON	Interviewer noted safety concerns about segment during segment listing or updating procedure (yes/no)
PHYSIMPED	Interviewer observed physical impediments to entry, such as locked door, community gate, etc. (yes/no)
URBAN	Address in an urban location (yes/no)
NCALLINFOLD	Total number of calls made to address
NCONTACT	Total number of contacts with household
SCR_SINGLEHH	Screener interview data indicates single person household (yes/no)
CHILD_LT15	Interviewer assessment before screener of whether anyone in household under 15 years of age (yes/no)
SEXUALLY_ACTIVE	Interviewer assessment before screener of whether anyone in household is married or cohabiting (yes/no)
AGE	Selected person age
HISPANIC	Selected person ethnic origin (Hispanic, not Hispanic)
BLACK	Selected person race (black, not black)
IWER_NONWHITE	Interviewer nonwhite (yes/no)
REGION	Geographic region (Northeast, Midwest, South, West)
BILINGUAL	Interviewer speaks English and Spanish (yes/no)
PSU_TYPE	Type of PSU (super eight, other large metropolitan, and smaller metropolitan or nonmetropolitan)
DOMAIN	Segment has in Census 2000 1) more than 10% population black, 2) more than 10% Hispanic, 3) more than 10% black and Hispanic, or 4) all other
QUARTER	Calendar quarter of data collection (1 to 16, where quarter 1 is approximately July 1 to September 30, 2006, quarter 2 October 1 to December 31, 2006, etc.)
MAINPROB	Interviewer assessment of probability (high, medium, low) of obtaining main interview made at week 7 of each quarter
BOTH_PHASE2	Address selected for Phase 2 sample as a main interview case (yes/no)

Estimating equation for totals

The NSFG estimator of the number of men or women with a particular characteristic can be computed as

$$\hat{N} = \sum_i w_{5i} I_i$$

where w_{5i} is the final trimmed weight for the i^{th} sample person and $I_i = 1$ if the person has the characteristic and $I_i = 0$ otherwise.

To estimate the total number of events or the total for a particular variable, such as the total number of live births, for persons in a particular subgroup identified by the indicator I_i is computed as

$$\hat{Y} = \sum_i w_{5i} I_i y_i$$

where y_i is the value of the characteristic of interest. The variable y_i may be a count such as number of live births or a continuous measure such as income.

Estimates of means or proportions can be constructed from these estimated totals. For example, the proportion of the population with a characteristic of interest can be computed as the ratio estimator

$$p = \frac{\sum_i w_{5i} I_i}{\sum_i w_{5i}}$$

Alternatively, the mean for a variable y_i can be computed as the ratio estimator

$$\bar{y} = \frac{\sum_i w_{5i} I_i y_i}{\sum_i w_{5i} I_i}$$

Sampling variances for these estimators must be computed taking the stratified multistage sample design into account. Variance estimation procedures are described in the last section.

Item Imputation

Overview of Item Imputation

In any survey, not every question is answered by every person interviewed. Sometimes a respondent cannot remember a fact asked for in a question; sometimes he or she refuses to answer. Other times, the answer that the respondent gives is clearly inconsistent with other information in the interview; one or more of the inconsistent answers is then set to ‘‘missing.’’ Such ‘‘missing data’’ create inconsistencies in estimates, which may be confusing for many users of the data. Assigning predicted values to these missing items is called ‘‘imputation.’’

Imputation has several advantages. It makes the data more complete and easier to use, eliminating the sometimes confusing decreases in sample size when cases with missing values are dropped from an analysis. It also allows all of the collected data to be used in analysis. In an analysis involving several different variables simultaneously, entire cases are dropped because they are missing for one variable, even though there are nonmissing values for the other variables for the case. The analyst thus discards a great deal of collected data by deleting cases with item-missing values. Discarding cases with missing values *implicitly assigns a value* to the missing items. Effectively, discarding cases assigns the average of the nonmissing values to each of the missing values. Imputation is a procedure that attempts to improve on this assignment by assigning a replacement value for each item-missing value that is a prediction from other variables, and not just an average of the same variable from cases without missing values.

Imputation is essentially based on a statistical (usually regression) model. Some analysts would prefer to use a different model than the one used for imputation. For example, an analyst may believe that a more statistically sophisticated model than the one employed in the imputation of the public-use data would provide a better fit and thus better imputed values.

Other analysts are concerned that standard errors using imputed values may underestimate standard errors of estimates. Thus, these statisticians urge analysts to use recently available methods for computation that account for imputed values in variance estimation.

Finally, there is a view that imputation is somehow a ‘‘fabrication’’ of data, and therefore that imputed values should not be used. However, imputation predicts values using the same methods analysts use to develop statistical models that summarize the associations present in the data. Given the low levels of imputation on most imputed values in the NSFG, however, these concerns are not likely to have any practical effect on most NSFG analyses, and they still have the advantages listed above—increasing sample size, preserving other reported data, and reducing bias.

There are thousands of variables in the NSFG data files. Of these, approximately 600 variables were selected for imputation because they are used frequently in analysis. These variables are referred to here as ‘‘recodes’’ or ‘‘recoded variables’’; some surveys use terms such as ‘‘constructed variables’’ or ‘‘generated variables.’’ Missing data for these recodes could create inconsistencies among survey estimates and confusion among data users about both the published data and the microdata file. Selecting, editing, and imputing this more limited set of variables was a way to ensure high-quality data for the variables used most often by NCHS and many other data users. Release of the data file could have been delayed significantly if missing data for all variables in the data file had been imputed.

The frequency of missing values for the recoded variables in the NSFG data was, as in previous NSFGs, low—in part because CAPI requires the interviewer to enter an acceptable response and then goes automatically to the next appropriate question. The CAPI program performs range and consistency checks to help prevent logically impossible answers. Typically no more than 1–2 percent of the values of all recoded variables, in male, female, and pregnancy files combined, are missing and subsequently being imputed.

The two imputation techniques used in the NSFG are:

- Logical imputation
- Regression imputation

Logical imputation involves having a substantive expert (usually at NCHS) look at a missing value, examine related variables, and assign a value to the missing value that is essentially an educated prediction of the true value.

Regression imputation, as used for the NSFG, uses software that imputes a missing value using potentially all other variables in the data set as predictors. In practice, logical imputation is rarely used in the NSFG. The vast majority of imputed values are imputed by regression imputation.

A major part of the work of imputation involves making certain that the values imputed are within acceptable ranges, and are logically consistent with other data reported by the respondent. Except when a reported value is obviously incorrect, actual reported data are never replaced by an imputed value. For each recoded variable in the database, an imputation flag identifies whether the value of that variable is imputed or not. Using the imputation flag, a researcher can identify the observations with an imputed value and the specific type of imputation procedure used for each specific recoded variable.

Detailed Item Imputation Procedures

In every survey some values are missing for some variables. These missing values create inconsistencies in reported results, since an analysis using different variables with different numbers of missing values will typically have unequal sample sizes. Item-missing values may also contribute to bias if the analysis is restricted to complete cases, those with valid values. If missing values differ on average from those observed in the valid cases, an analysis that only uses cases with complete data may yield biased estimates.

There are several remedies for item-missing values, including weighting specific to each analysis, and imputation—replacing missing values with predicted values (see reference 18). The latter, imputation, makes analyses more consistent and can reduce bias due to item nonresponse, depending on the predictive power of the imputation procedure.

As in previous NSFG cycles, imputation was used only for a subset of variables, those to be used in the primary tabulations and analyses released by NCHS. Three files required imputation for some variables: female, male, and pregnancy. Each has somewhat different sets of variables. There are approximately 600 variables (called ''recodes'' in the NSFG data files) identified for imputation across the three data files in the NSFG.

For all variables with imputed values, a set of indicator variables was generated in the female, male, and pregnancy files that designated which cases received imputation for each variable. These ''imputation flags'' may be used by the analyst to replace the imputed values in the data set with values predicted from a model or procedure specified by the user.

There is a considerable range of methods used to impute for item-missing values in survey practice (18,19). In some instances, values can be replaced through a logical derivation based on the relationship among variables. In most cases, though, imputation must be based on a form of prediction. The imputation procedures used in survey practice range from simple cell mean procedures to sequential hot deck, flexible matching, and regression imputation (19).

Previous NSFG cycles have used a combination of logical, hot deck, and regression imputation procedures. In the 2006–2010 NSFG, most imputations are being performed using a regression procedure. The procedure is programmed to impute all recoded variables within a data file. A large number of potential predictors was used in the regression procedure. The potential predictors for any given recoded variable consist of all recoded variables and all variables used to ''construct'' recoded variables. (In addition, certain logical constraints were built in when necessary.) The imputation procedure itself employs a generalized linear regression model, providing for the imputation of interval, dichotomous, ordinal, nominal, and count scales for the variable being predicted.

Sequential regression

Regression imputation in the NSFG is implemented using a sequential regression method described in Raghunathan et al (20). The method employs both a sequential procedure to impute multiple variables in a particular order and Bayesian methods to perturb regression coefficients used in predicting values and to add residuals to imputed values. The method allows for alternative scales for each variable to be imputed, including interval, dichotomous, ordinal, nominal, and count scales.

For several scales, predicted values outside the range of acceptable values could be generated. For example, for dichotomous scales, under logistic regression, predicted proportions between zero and one are generated. These values were converted to zero or one values through a random process. For example, suppose that the predicted value for a dichotomous variable is 0.4. A uniform random number from zero to one is generated. If the random number is greater than 0.4, the imputed value of ''1'' is assigned. Otherwise, the imputed value of ''zero'' is assigned. Similar random assignment of valid values occurs for polytomous and count variables.

Imputed values from other variables are used in the prediction of values used as imputations for a given variable. The procedure is multivariate, using all variables specified as predictors of each variable in the imputation procedure, and thus effectively preserves existing covariance structure in the data.

Consider a set of $p + q$ variables from the NSFG, where the p variables do not require imputation (all values were observed, or logically imputed prior to regression imputation).

The imputation required several cycles. In the first, the first of the q variables, y_{p+1}, is regressed on all the p variables without missing values. A linear, logistic, multinomial, or Poisson regression is used, depending on the measurement scale for y_{p+1} (the variable being imputed) specified by the user. From this regression, a predicted value $\hat{y}_{p+1}$ is generated for each missing value. Still in the first cycle, the second variable, y_{p+2}, is regressed on the p variables without missing values and y_{p+1} (including its imputed values). Predicted values for y_{p+2} from this estimated regression model are assigned to item-missing values. This cycle continues by repeating the regression imputation process for the variables $y_{p+3}, ..., y_{p+q}$, imputing in each case using all p variables with no missing values and all imputed variables obtained to that point.

Cycles two, three, and so on (a total of five cycles are typically used) repeat the regression imputation, but include all variables, the p with no missing values and the other q–1 with imputed values, as predictors. Thus, y_{p+1} is regressed on the set of predictors $y_1, ..., y_p, y_{p+2}, ..., y_{p+q}$, and imputed values obtained from the predicted values under the regression model.

The predicted values from each regression in each cycle are obtained through a process that perturbs the coefficients and adds a random residual. The process is described in the appendix of Raghunathan et al. (20).

Variables may have restrictions to limit the set of cases to be imputed. For example, persons who have never had sex should not receive an imputed value for age at first intercourse. Restrictions are readily programmed in the software. Values to be imputed may be bounded within specified limits. Bounds for imputed values can also be set and truncated distributions are used to assign predicted values. In the NSFG, a major part of the work of imputation is in specifying appropriate bounds for imputed values because of the extensive inter-relationships within the data.

The sequential regression imputation method is implemented in the IVEware software system http://www.isr.umich.edu/src/smp/ive/.

Variance Estimation

Overview of Variance Estimation

Sampling variance is a measure of the quality of a statistic (such as a proportion or a mean) caused by having taken a sample instead of interviewing the full population. For example, in the 2006–2010 NSFG, the sampling variance measures variation caused by interviewing the continuous NSFG sample instead of the full population—approximately 125 million women and men 15–44 years of age in the entire country. Sampling variance measures the variation of the estimated statistic over repeated samples. The sampling variance is zero when the full population is observed, as in a census.

For the NSFG, the sampling variance estimate is determined by sampling design and the population parameter being estimated, and it is called the design-based sampling variance. The NSFG data files contain a final weight and information necessary to estimate the sampling variance for a statistic. Many statistical software packages by default compute "population" variances, which may severely underestimate the sampling variances.

Special software is required to accurately estimate sampling errors in a complex sample such as the NSFG, but such software is increasingly common and easy to use and obtain. For example, by default, SAS produces population variances, but it has Taylor series expansion and repeated replication procedures for complex survey estimates in specialized "SURVEY" procedures. 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 are on the NSFG website at http://www.cdc.gov/nchs/data/series/sr_02/sr02_142.pdf (accessed February 6, 2009; also available in reference 21). In addition, examples of variance estimates for the 2006–2010 NSFG will be presented with documentation for each data release.

Detailed Description of Variance Estimation

Since NSFG estimates are based on a sample rather than a complete enumeration of the eligible population, they are subject to sampling error, a departure between the true population value and the estimated value. This difference may be due to systematic or fixed sources of error, such as nonresponse or noncoverage. The difference is also attributable to variable sources of error, including the use of a sample to represent the population.

Probability sampling, as used in the NSFG, allows the direct estimation of one of these sources of error, the variable error due to sampling. A considerable share of the survey design and estimation literature develops proper procedures for estimating the sampling variance under different sample selection techniques. See Wolter (22) for a review of variance estimation techniques.

The estimation of sampling variance for NSFG estimates requires procedures that properly account for the principal effects of the different sampling techniques employed in the NSFG. There are four principal design features that can be accounted for:

- Stratification of PSUs.
- Selection of PSUs (cluster sample selection).
- Weights that adjust for unequal probabilities of selection and other design features.
- The presence of imputed values.

One way to handle imputed values in variance estimation is with multiple imputation. Multiple imputation is a process with three complementary parts. The first part of multiple imputation is that several (two or more) imputed or predicted values are generated for a missing item for each case with missing values. Second, estimates are produced by computing an estimate using each of the imputed values. For example, suppose that five imputed values are generated for a particular variable for each case with a missing value for the variable. Then an estimate is computed using the set of first imputed values and the remaining valid or real values. Another estimate is computed using the second imputed values and the valid or real values. Five estimates in total are computed. The final estimate is the average of the five estimates. The third part is that the estimate is now subject to variability from two sources, the sample design used to select the overall sample, and the imputation process used to generate imputed values.

To capture these variance sources properly, the variance of each of the five estimates is computed using the imputed and valid values, taking into account the complex sample design. The five variance values are averaged. In addition, the variance among the five estimates is computed. The total variance of the overall estimate (the average of the five estimates, one per set of imputed values) is the sum of the average variance of the five estimates plus the variance among the five estimates. This approach captures fully both sets of variability.

These methods are becoming increasingly more accessible in existing statistical estimation software. However, to implement the method, these software systems require that there be a data set for each set of imputed values. If there are five imputed values, five data sets are needed. Because of the size of the NSFG data files, the computational burden on users is potentially prohibitive, even with the availability of procedures in software. While the IVEware software used for the NSFG has the capacity to generate multiply-imputed data, for public release purposes a single imputed data set, with a single imputed value for imputed variables, is being released.

The remainder of this section considers the effect of the NSFG design (including stratification, oversampling, and weighting) on estimated sampling variances, the availability of software that can estimate variances taking these design features into account, and the creation of codes that can be used for variance estimation.

If software is used that does not account for the complex design, sampling variances will be under-estimated. The consequence will be test statistics that generate smaller probability test values than are actually correct, and confidence intervals that are narrower than they should be. Both of these types of errors can lead to incorrect inferences. For example, smaller probability test levels might lead an analyst to reject a null hypothesis incorrectly.

For this reason, all users of NSFG data should use variance estimation procedures discussed in this section. They will assure that incorrect test procedures or overly narrow confidence intervals are not used in drawing conclusions from the data.

Summary of Variance Estimation Principles

There are two primary approaches used to estimate sampling variance from complex sample surveys: Taylor series approximation methods and repeated replication methods. The former are based on an analytic treatment of statistical estimates, while the latter are based on computer intensive resampling of the survey data.

Complex sample survey data generate statistics which, through the use of weights, can be estimated by standard statistical software. That is, the estimated percentages, means, regression coefficients, and other statistics are computed properly using weighted data. However, for the purpose of estimating test statistics and standard errors of estimates, standard software treats the data as though the sample was selected using very simple sampling methods. The consequence, as noted previously, is an underestimation of standard errors, and an overstatement of the significance of test statistics.

The Taylor series approximation, or linearization procedure, can be adapted to account for the complex design features. Taylor series linearization is a widely used technique in statistics for obtaining approximate variance estimates for nonlinear statistics. The Taylor series is an infinite series expansion of a function that has a valuable property: for ‘‘well behaved’’ functions, the successive terms in the expansion rapidly approach zero. Thus, for some functions, the first term of the Taylor series expansion can be used to represent the entire infinite series.

The first term is linear in the variables in the function. Variances are computed for the ‘‘linearized’’ function. The properties of the Taylor series expansion approximation variance estimates are well understood empirically (see reference 23 for evidence from complex sample surveys). It has been applied to a wide range of statistics computed from sample survey data, from means and proportions to logistic regression coefficients estimated using iterative numerical solutions.

For NSFG data, the Taylor series method can be adapted to the sampling error computing units (SECUs) created to represent the stratification and clustering in the sample. These variance estimation procedures are widely implemented in software systems that implement Taylor series approximations for complex sample survey designs.

The repeated replication procedures are based on early work of Mahalanobis (24) and Deming (25). They proposed and encouraged the use of interpenetrating or replicating sampling methods that simplified variance estimation considerably. For example, if a sample were comprised of T replicate samples, each selected in the same way, and a statistic $\hat{\theta}_t$ was computed from each replicate, the sampling variance of the average

$$\bar{\theta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\theta}_t$$

can be computed simply as

$$var\,(\bar{\theta}) = \frac{1}{T(T-1)}\sum_{t=1}^{T}(\hat{\theta}_t - \bar{\theta})^2.$$

The idea of replicated sampling was extended to survey data by McCarthy (26), who proposed selecting replicate samples from a larger sample by employing in the replicates selections that accounted for principle sampling methods used in the original sample selection. More appropriately for NSFG data, a ‘‘jackknife’’ replication procedure ‘‘drops one’’ SECU in a stratum, and retains all other SECUs and their cases in the sample. In the NSFG, the cases in the SECU remaining in the stratum from which the one SECU was dropped must have their values increased (doubled) to account for the under-representation of the stratum.

With repeated replication, the sampling variance for a statistic can be estimated in a form very similar to that given previously for replicated sampling. Let $\tilde{\theta}$ denote the estimate of the statistic of interest computed from the entire sample. Then,

$$var(\tilde{\theta}) = \frac{1}{T}\sum_{t=1}^{T}(\hat{\theta}_t - \tilde{\theta})^2,$$

where $\hat{\theta}_t$ is computed from the t^{th} replicate. For the jackknife repeated replication procedure,

$$var(\tilde{\theta}) = \sum_{t=1}^{T}(\hat{\theta}_t - \tilde{\theta})^2.$$

See Rust (27) for a more detailed review of these procedures.

Variance Estimation Software

These variance estimation procedures are all available in a number of commercially available computer software packages, or for free download over the Internet in a few instances. In earlier NSFG cycles, it was often complex and difficult to obtain and use variance estimation software, but several software packages now allow users to estimate variances for means, proportions, regression coefficients, logistic regression coefficients, and other statistics directly within the larger software system. These statistical software packages implement the Taylor series approximation, repeated replication methods, or both approaches for a complex sample survey like the NSFG.

There are two basic types of software systems available for this purpose: stand-alone and integrated packages. The stand-alone software packages require users to input data into a special format used by the system. Integrated software for estimation from complex sample survey data allows a user to conduct an analysis without converting data to another format. The user can construct a survey data set in a format used by a major statistical software package, such as SAS, Stata, or SPSS, and then compute estimates and test statistics within those systems that take complex design features into account. For example, SAS version 10 has four PROCedures which properly account for complex features such as stratification, clustering, and weighting. SUDAAN is a stand-alone system that comes in a SAS compatible version that allows SUDAAN commands to be embedded in SAS programs as though they were part of the SAS language. Users with SAS data sets can thus analyze data using either SAS or SUDAAN, although they will find more analysis procedures available in SUDAAN. Stata has a larger set of survey or ''svy'' versions of frequently used analysis commands that account for stratification, clustering, and weighting in complex sample surveys. The SPSS Complex Sample module that can be added to the base statistics module for complex sample surveys has features for frequency tables and several types of regression models.

NSFG data users can find descriptions of these and other software systems for estimation from complex sample survey data, along with detailed explanations of their features, at a website maintained by the Survey Research Methods Section of the American Statistical Association, currently located at http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html (28). NSFG data users are encouraged to visit this website and become more familiar with the features of these software systems.

Empirical comparisons of these software packages (28) have shown that they provide essentially identical estimates of statistics of interest, and that the estimated sampling variances of the estimated statistics are virtually the same, regardless of the computational method used to estimate the sampling variances. The choice among these systems can thus be based on practical considerations of convenience, cost of software acquisition and maintenance, and current data format.

Sampling Error Computing Units

The 2006–2010 NSFG sample has 110 PSUs. Each sample PSU was drawn from a stratum of one or more PSUs. Sampling variance cannot be estimated directly from this type of a design, since there is but one selection per stratum. The NSFG PSUs are arranged into a set of sampling error computing units (SECUs) that could be grouped into pairs for variance estimation purposes.

The 28 Self-Representing (SR) PSUs (the ''super 8'' plus 20 other metropolitan SR PSUs) are divided into two or four representative units each by a systematic sample of the segments within each PSU. For example, in a large SR PSU the segments are numbered in sample selection order within each sample domain as 1, 2, 3, 4, 1, 2, 3, 4, 1, 2 .... Segments with the same number are grouped together to form a pseudo PSU representing the entire MSA. The first and third such ''combined units'' are paired to form a stratum with two combined units, or SECUs. Thus, the large PSU in this example has four SECUs in the data set, grouped into two pairs. Each pair is a pseudo-stratum. Smaller SR PSUs are divided into two SECUs, which are in turn grouped into a single pair as a pseudo-stratum. This SR PSU process generates a total of 36 stratum pairs of SECUs.

For the remaining 80 Non-Self-Representing (NSR) PSUs, the strata are inspected to identify groups of four that were as alike as possible. A total of 20 sets of four PSUs were created for purposes of variance estimation.

In combination with the SR PSU SECUs, there are thus a total of 36 pairs of SECUs and 20 sets of four, or a total of 152 SECUs. Used in pairs and groups of four, the total degrees of freedom for variance estimation from the female, male, or pregnancy file is thus approximately 96. Two variables, SEST and SECU, are in the data set to identify these groupings. The codes for these values have been randomly ordered to mask the identity of the units.

The National Survey of Family Growth Continuous Design: A Guide for Analysts

If the NSFG plan for continuous interviewing (29) is carried out for a period of years, it will allow the release of the data more frequently than in the past, as it is collected.

Analyzing the Continuous National Survey of Family Growth as Periods of Data Collection

The present plan is that under the sample design described in this report, there would be two releases of NSFG 2006–2010 data.

- The first data file includes about 13,500 interviews collected during the 2.5-year (30-month) period from July 1, 2006–December 2008 (''the 2006–2008 data'').
- The second will include over 20,000 interviews collected during the 4-year (48-month) period from July 1, 2006, through June 30, 2010.

For each of these data releases, users can treat the data much as they might have handled previous NSFG cycles. That is, a user will analyze the 2006–2008 data as the next ‘‘cycle’’ of the NSFG. The structure of the 2006–2010 NSFG data is very similar to that of the 2002 NSFG, and there are weight and sampling error codes that allow the user to compute estimates for the entire 30-month period during which data were collected. Users can compute estimates and fit models as they did for the 2002 NSFG or earlier cycles.

The principal difference is simply that in previous cycles, the data were collected in 12 months or less, but in the current NSFG, the data were collected over a longer time period. In the 2002 NSFG, for example, data were collected over a 12-month period (March 2002–February 2003), and users referred to the findings as being about the U.S. population in 2002, using a single weight and a set of sampling error codes.

With the first release of the NSFG continuous design, a user could repeat an analysis and compare results for 2002 with results for the period mid-2006 through 2008. The 2006–2008 data set weight is ‘‘centered’’ in the middle of this time period. The population for which inferences can be made will thus be the average U.S. population over the period from approximately July 1, 2006, through December, 31, 2008, for the first release—as it was in 2002, but the time period is longer.

By interviewing about 5,500 respondents per year, the NSFG can give data users larger sample sizes than were available from earlier cycles (and obtained at a lower cost per completed interview), but the user must refer to periods of time longer than 1 year. The larger sample sizes make more detailed study of the entire sample or subgroups possible than before. The analysis is done the same way as in earlier cycles of the NSFG.

Questions added in 2007—Some new questions were added to the NSFG questionnaire in July 2007. The variables generated by these new questions are contained in the first (30 month) release, and were collected for an 18-month period from July 2007, through December 2008. Any analysis conducted that uses these 2007 variables will have missing or ‘‘inapplicable’’ codes for respondents interviewed before the questions were added to the survey. Sample sizes will be reduced accordingly, from approximately 13,500 to about 8,000. To compensate, a separate weight will be provided for the variables introduced in 2007 (‘‘2007 variables’’) so that NSFG data users can analyze them.

Analyzing the 2006–2010 National Survey of Family Growth for shorter periods of time

There are of course several shorter time periods within the expected NSFG releases that could be selected by a user. However, as the time period becomes shorter, such as moving to a single year or even a calendar quarter within a year (such as the period January through March 2007), the sample sizes become smaller, and estimates become less reliable. At a certain level, the NSFG estimates will be based on such small sample sizes that any estimates produced will no longer have adequate precision.

At the present time, NSFG staff are examining whether estimates for single calendar years will be adequately reliable for any purpose. A calendar year contains approximately 5,500 completed female and male interviews. Given the NSFG's oversampling of minority groups, teenagers, and women, and its clustered design, this sample size is unlikely to produce statistically reliable estimates for most purposes. For the anticipated first release of 2006–2008 data, which contains only the two calendar years 2007 and 2008, calendar year estimates are <u>not</u> recommended. No calendar year weights will be provided in the first data file release.

As a result, NSFG <u>does not recommend</u> that estimates be computed for intervals shorter than 18 months. Estimates for individual calendar years, or individual calendar quarters, should <u>not</u> be computed, and weights to allow users to compute such estimates will <u>not be provided</u>.

When the second and subsequent data files are released from continuous interviewing, other analytic options may be possible, including analyzing trends in 1 or 2-year estimates. Research is being conducted to determine whether and under what conditions such estimates will be possible. If they are determined to be feasible, weights will be provided for them in a future data release.

Once several calendar-year time periods of NSFG data become available, users may then compare estimates across time. Guidance on how to make comparisons of these time periods will be provided in the user guide for the 2006–2010 release. Again, precision for comparisons will be limited, due to the smaller sample sizes, and differences will have to be quite large to be statistically significant.

Analyzing the 2006–2010 National Survey of Family Growth for small subgroups or rare behaviors

The continuous NSFG allows users the opportunity to accumulate data across several years of data collection to study small subgroups or rare behaviors. Accumulation is useful when a short time period simply does not provide enough cases for adequate levels of precision in estimates. In the expected 2006–2010 release, users will (with the overall larger sample size) be able to select proportionately smaller subgroups for analysis than possible with previous NSFG cycles.

Yet, there will be limits to the size of sample that should be analyzed. Guidelines regarding the minimally acceptable levels of precision for such subgroup analyses have not yet been determined. However, NSFG staff themselves exercise caution in publishing or releasing findings for estimates of prevalence or averages whenever a standard error is more than 25 percent of the size of the estimate. Estimated proportions or means where the standard error is larger than 25 percent of the estimate itself, or where the numerator or denominator are too small, must be interpreted with caution.

Frequently Asked Questions: Summary in Question-Answer Format

To summarize some of the practical issues of interest to users of the data file, the following section shows some Frequently Asked Questions (FAQ's) and their answers. Additional information will be released on the NSFG website http://www.cdc.gov/nchs/nsfg.htm as it becomes available.

1. **Given that the 2006–2010 design uses quarterly samples, can I analyze the data for just one quarter?**

   No. Sample sizes for a single quarter are too small to provide estimates with adequate levels of precision. In the first data file, weights will be included only for the full 30 months of interviewing, and for the last 18 months of interviewing (when certain new questions were introduced). Analysts should only use time periods for which sampling weights are provided.

2. **The public-use data file has both recoded variables (''recodes'') and raw variables. Which ones should I use?**

   NCHS recommends using recodes, when they are appropriate for a given analysis, for two main reasons. First, the recoded variables have been studied for consistency and any missing data have been resolved through imputation. Thus, using the recodes allows the analyst to make use of the intensive scrutiny given to them. Second, NCHS uses the recoded variables in many of the tables in NCHS official publications, so users can check their own results with those of NCHS to ensure comparability. If a recoded variable is not available for a specific analysis, then the raw variables or computed variables should be used.

3. **Were different questionnaires used over the 4 years of data collection (2006–2010)?**

   Changes occurred at the beginning of the 2nd and 3rd years of interviewing (no changes were made for year 4). But over 95 percent of the questions remain the same from year 1 to year 4 of interviewing. Variables measured in only some of the years are noted in the public-use documentation of the data sets.

4. **How do I combine the different quarters of data collection in my analysis?**

   Sampling weights will be provided on the data files for time periods for which analysis is appropriate. See the data release documentation for full guidance on your specific analysis.

5. **Given that the size of the data set can become large over several years of interviewing, can I analyze the data separately for different states in the country?**

   Although the number of data records in the survey can become large when several years of data are combined, the sample is limited to 110 PSUs or areas. These primary areas do not fall in all states. As long as this design is used, estimates cannot be computed for individual states.

6. **Can I combine the data for males and females?**

   Yes. The data files contain some of the same variables for males and females. Using the sampling weights, estimates for males and females combined can be made.

7. **For a recoded variable, how do I find out what questions in the questionnaire contributed to it?**

   The ''recode specifications'' are given in the public-use data file documentation. These specifications show how each recoded variable was constructed. The NSFG staff recommends using recodes when they are appropriate for a particular analysis, because they have been checked for accuracy, edited for consistency, and imputed.

8. **Should I obtain the same results on birth-related statistics from analyzing the NSFG as vital statistics?**

   Birth statistics based on the National Vital Statistics System are derived from a complete count of the approximately 4 million birth certificates filed each year. The NSFG, in contrast, estimates these births with a sample of a few hundred births each year to women in the continuous NSFG sample. Therefore, NSFG estimates will not match those from the birth certificates exactly, primarily because of sampling error. But there are also some differences between the coverage, nonresponse, and measurement features of the two approaches to estimation.

9. **What format is used for the public-use data set?**

   The data sets are released in ASCII format and are compatible with several major statistical software systems that permit analysis of complex survey data, including SAS, SPSS, and Stata.

10. **How is the data set any different from analyzing any other sample survey?**

    For the most part, analyzing the NSFG data is no different than analyzing a previous NSFG cycle, at least in terms of the application of standard survey estimation software to the data file. Each of the data files released from the 2006–2010 NSFG data can be thought of and treated as a single cycle of the NSFG, but the time period to which the results apply must be described accurately.

    Previous NSFG cycles allowed the user to report estimates for a single year. The continuous NSFG will allow the user to report estimates for a period of a few consecutive years. For example, the proportion of women 15–44 years of age who are currently using the oral contraceptive pill may be reported for the period from 2006 to 2008. Once the 2009–2010 data are released, the same statistic can be computed for 2006–2008, 2009–2010, or 2006–2010. The third period (2006–2010) allows the user to report the rate across a longer time interval, but with greater precision, since the sample size of the pooled data will be larger.

In all these estimates, just as in previous NSFG files, the analyst must use appropriate weights and variance estimates. The appropriate weights and sampling error codes will be provided with each release of the NSFG that allow proper estimation for different time periods of interest. The presence of alternative time period weights will require users to choose the time period of estimation and the appropriate weight to apply, but the benefits include greater analytical flexibility and larger sample size.

References

1. French DK. National Survey of Family Growth, Cycle 1: Sample design, estimation procedures, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(76). 1978.
2. Freedman R, Whelpton PK, Campbell AA. Family planning sterility and population growth. New York: McGraw-Hill. 1959.
3. Whelpton PK, Campbell AA, Patterson JE. Fertility and family planning in the United States. Princeton, NJ: Princeton University Press. 1966.
4. Ryder NB, Westoff CF. Reproduction in the United States, 1965. Princeton, NJ: Princeton University Press. 1971.
5. Westoff, CF, Ryder NB. The contraceptive revolution. Princeton, NJ: Princeton University Press. 1977.
6. Grady, WR. National Survey of Family Growth, Cycle 2: Sample design, estimation procedures, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(87). 1981.
7. Bachrach CA, Horn MC, Mosher WD, Shimizu I. National Survey of Family Growth, Cycle 3: Sample design, weighting, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(98). 1985.
8. Judkins DR, Mosher WD, Botman S. National Survey of Family Growth: Design, estimation, and inference. National Center for Health Statistics. Vital Health Stat 2(109). 1991.
9. Potter FJ, Iannacchione VG, Mosher WD, Mason RE, Kavee RD. Sample design, sampling weights, imputation, and variance estimation in the 1995 National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 2(124). 1998.
10. Kelly JE, Mosher WD, Duffer Jr. AP, Kinsey SH. Plan and operation of the 1995 National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 1(36). 1997.
11. Groves RM, Benson G, Mosher WD, Rosenbaum J, Granda P, Axinn W, Lepkowski JM, Chandra A. Plan and operation of Cycle 6 of the National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 1(42). 2005.
12. Kish, L. Survey sampling. New York: J. W. Wiley and Sons. 1965.
13. Kish L, Scott A. Retaining units after changing strata and probabilities. J Am Stat Assoc 66:461–70. 1971.
14. Hansen MH, Hurwitz WN. The problem of nonresponse in sample surveys. J Am Stat Assoc 41:517–29. 1946.
15. Groves RM, Heeringa S. Responsive design for household surveys: Tools for actively controlling survey errors and costs. J Royal Stat Soc A 439–57. 2006.
16. Groves RM, Couper MP. Nonresponse in household surveys. New York: John Wiley and Sons. 1998.
17. American Association for Public Opinion Research. Standard definitions: Final dispositions of case codes and outcome rates for surveys. Available from: http://www.aapor.org On-line edition 3.1. Revised 2008.
18. Groves RM, Dillman DA, Eltinge JL, Little RJA, editors. Survey non-response, New York: Wiley and Sons. 2001.
19. Kalton G, Kasprzyk D. The treatment of missing survey data. Survey Methodology 12:1–16. 1986.
20. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. Sequential regression imputation for survey data. Survey Methodology 27:85–96. 2001.
21. Lepkowski JM, Mosher WD, Davis KE, et al. National Survey of Family Growth, Cycle 6: Sample design, weighting, imputation, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(142). 2006.
22. Wolter KM. Introduction to variance estimation. New York: Springer-Verlag. 1985.
23. Kish L, Frankel MR. Inference from complex samples. J Royal Stat Soc B 36:1–37. 1974.
24. Mahalanobis PC. On large-scale sample surveys, Philosophical transactions of the Royal Soc B 231:329–451. 1944.
25. Deming WE. On simplifications of sampling design through replication with equal probabilities and without stages. J Am Stat Assoc 51:24–53. 1956.
26. McCarthy PJ. Replication. An approach to the analysis of data from complex sample surveys. National Center for Health Statistics. Vital Health Stat 2(14). Washington, DC: U.S. Government Printing Office. 1966.
27. Rust K. Variance estimation for complex estimators in sample surveys. J Official Stat 1:381–97. 1985.
28. Survey Research Methods Section, American Statistical Association. Summary of survey analysis software. Available from: http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html. Accessed July 7, 2009.
29. Groves RM, Mosher WD, Lepkowski JM, Kirgis NG. Planning and Development of the Continuous National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 1(48). 2009.

Appendix I. Glossary

Audio computer-assisted self-interviewing (ACASI)—An interviewing technique in which the respondent uses a laptop computer to complete a questionnaire. The respondent uses earphones, which deliver an audio recording of the questions, and reads the question text on the laptop monitor. The respondent chooses a desired response option to each question, using the laptop's keyboard. The software directs the respondent to the next appropriate question based on the answers entered. The respondent performs these steps out of the sight of the interviewer, in an attempt to offer the respondent as much privacy as possible. ACASI is offered in both English and Spanish in the continuous NSFG.

Blaise®—A software system that presents the questions in a questionnaire, such as the NSFG. Blaise® is programmed to route the respondent to the next appropriate question, store the respondent's answers, and check the consistency of one answer with answers to other related questions. Blaise® was used in the 1995, 2002, and 2006–2010 NSFG.

Computer-assisted personal interviewing (CAPI)—An interviewing technique in which the interviewer uses a laptop computer in the interview. The laptop displays question text for the interviewer to read and provides any other necessary instructions to the interviewer. Interviewers record the respondent's answers using the keyboard. Software directs the interviewer to the next appropriate question based on the answers entered.

Contact rate—At the screener stage, the contact rate is the percentage of sample households where an interviewer talked with someone at the household (i.e., the screener contact rate). At the main interview stage, the contact rate is the percentage of sample persons who met with the interviewer on one or more visits to the household by the interviewer (i.e., the main interview contact rate).

Cooperation rate—The percentage of sample households that were contacted and granted a screener interview (i.e., screener cooperation rate); or the percentage of sample persons contacted who granted a main interview (i.e., main interview cooperation rate).

Coverage error—Deviations between the characteristics (for example, values of estimated population characteristics) of the sampling frame and the desired target population. Coverage errors arise from the failure to include some households containing eligible persons in the list of households within segments and failure to list some eligible persons within sample households on the sampling frame.

Delivery Sequence File (DSF)—The DSF from the U.S. Postal Service lists all addresses to which mail is currently delivered by the Postal Service. In most areas, the DSF is the basis for a list of housing units from which listing for the NSFG is done.

Domain—A stratum; a group of sampling units (such as blocks) placed in the same subset from which a sample of units was selected.

Double (or two-phase) sample—A subsample of nonrespondent sample cases selected after the completion of a phase of data collection. NSFG used such a subsample in Cycle 6 (2002) and in 2006–2010.

Eligible household—A household containing at least one person who was eligible for the NSFG—that is, males or females 15–44 years of age at the date on which the screener was completed, and living in the household population of the United States (all 50 states and the District of Columbia). Whether a selected household has an eligible person is not known until the household screener is conducted. If a household has two or more persons 15–44 years of age, one person is selected randomly.

Eligibility rate—The percentage of sample cases that are members of the target population. In the NSFG the eligibility rate is the percentage of households that contain a person aged 15–44.

Equal probability selection method (Epsem)—A sample design that gives all sample units an equal chance of selection.

Institute for Social Research, University of Michigan—The Institute for Social Research (ISR) at the University of Michigan conducts the fieldwork and data processing for the 2006–2010 National Survey of Family Growth (NSFG) under a contract with NCHS. ISR has several centers that participated in the NSFG: the Survey Research Center provides overall coordination and is responsible for data collection, weighting, and variance estimation; the Interuniversity Consortium for Political and Social Research processes data and develops documentation and web-based systems; and the Population Studies Center provided substantive expertise on demography and family growth.

Institutional Review Board (IRB)—A committee of peer and community reviewers of research procedures involving human subjects that weighs the benefits of the research relative to the risks of harm to human subjects. The NSFG was reviewed and approved by the NCHS IRB, which NCHS calls a ''Research Ethics Review Board,'' or RERB.

Item imputation—The process of assigning answers to cases with missing data (''don't know,'' ''refused,'' or ''not ascertained''). In the NSFG, item imputation is only performed on approximately 600 ''recoded variables,'' or ''recodes'' (defined in the following text, under ''recodes''), rather than all of the thousands of variables in the data set. The purposes of imputation are to make the data more complete, more consistent, easier to use, and, most importantly, to reduce bias caused by differential failure to respond. For example, if a respondent's educational level is missing and a value of ''high school graduate'' is assigned, education is imputed. Imputation is done in two ways in the 2006–2010 NSFG, logical and regression imputation. Regression imputation uses a regression equation to estimate a value for a case with missing data. Regression imputation was used to assign most of the imputed values. Occasionally, however, logical imputation is used: logical imputation uses a subject-matter expert to assign a value based on the value of other variables for the case with missing data.

For nearly all of the recoded variables for which imputation is done in the continuous NSFG, less than 2 percent of the cases received an imputed value.

Life history calendar—A visual presentation of a calendar covering the reference period of various questions. A life history calendar is used to help the respondent record the dates of events, which are used as ''anchors'' to cue memories of the dates of events measured in the survey. In the 2006–2010 NSFG, the female interview used a life history calendar as a recall aid for the pregnancy and contraceptive history portion of the interview.

Main interview—An interview sought within sample households containing an eligible target population member. If the screening interview reveals that the household contains one or more persons 15–44 years of age, a main interview is requested from one of those persons. If there are two or more persons 15–44 years of age, one such person was selected at random for the main interview.

Measure of size—A value assigned to every sampling unit in a sample selection. Typically measures of size are a count of units associated with the elements to be selected. For example, measures of size for NSFG PSUs are the count of occupied housing units obtained in the 2000 Census of Population and Housing, since sample selection with the PSU would have selected housing units.

Measures of size are also used in the selection of eligible persons within the household (see Figure 5) to increase the chances of selection of such groups as teenagers 15–19 years of age, black and Hispanic persons, and females. Each person in the household is assigned a measure of size between zero and one, where the measures are predetermined values for each age by gender by race and ethnicity group. The measures of size are cumulated across eligible persons, a random number from zero to the sum of the measures generated, and an individual selected based on the cumulated measures of size.

Multiphase design—A survey design that changes its sample design or recruitment protocol over different sets of sample cases or over time periods of the survey to obtain an optimal balance of costs and quality of survey estimates.

National Center for Health Statistics (NCHS)—NCHS is the Nation's principal health statistics agency. It designs, develops, and maintains a number of systems that produce data related to demographic and health concerns. These include data on registered births and deaths collected through the National Vital Statistics System, the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), the National Health Care Surveys, and the National Survey of Family Growth (NSFG), among others. NCHS has conducted the NSFG since 1973. NCHS is one of the ''Centers'' for Disease Control and Prevention (CDC), which is part of the U.S. Department of Health and Human Services.

Office of Management and Budget (OMB) Clearance—OMB reviews survey materials and questionnaires proposed for use by government agencies under the provisions of the Paperwork Reduction Act. The review is conducted by the OMB's Office of Information and Regulatory Affairs. The NSFG was reviewed by this office of OMB.

Paradata—Information collected via computer software or interviewer observations describing the sample unit, interactions with sample household members, or features of the interview situation. The NSFG used observations of characteristics of sample housing units to reduce the number of callbacks; used statements made by household screener informants to diagnose their concerns about the survey; used call record data to model the probability of obtaining an interview on the next visit; and used observations of the respondent during ACASI for measurement error modeling. Some paradata are labeled as ''process data.''

Phase—A period of data collection during which the same set of sampling frame, mode of data collection, sample design, recruitment protocols, and measurement conditions are used. In the 2006–2010 NSFG there are two phases in each 12-week quarter: first, in weeks 1–10, the standard protocol is used, although paradata are used to optimize the efficiency of the interviewers; second, in weeks 11–12, a subsample of nonrespondents from Phase 1 is offered higher incentives and certain other rules are changed. (See text for details.)

Public-use file—An electronic data set containing respondent records from a survey with a subset of variables collected in the survey that have been reviewed by analysts within NCHS to assure that the identities of the respondents are protected. This file is disseminated by NCHS to encourage widespread use of the survey.

Primary sampling unit (PSU)—The first-stage selection unit in a multistage area probability sample. In the NSFG, PSUs are counties or groups of counties in the United States; there were 110 PSUs selected into the NSFG sample for 2006–2010.

Race and ethnicity—Race and ethnicity is used in this report as it was used to design and select the NSFG sample. Three categories were used: Hispanic, non-Hispanic black, and all other. Hispanic and non-Hispanic black men and women are selected at higher rates than others in the NSFG to obtain adequate numbers of Hispanic and black persons to make reliable national estimates for these groups. Thus, in this report, tables showing ''race and ethnicity'' show the three categories used to design and select the sample. In contrast, in reports that are designed to present substantive results, the ''all other'' category is often split into ''non-Hispanic white'' and ''non-Hispanic other'' categories.

Recodes or recoded variables—It is not possible to edit or impute all of the variables in the continuous NSFG data file. NSFG staff selected about 600 variables from the NSFG data file that were constructed, edited, and imputed. These are called recodes or recoded variables. Recodes are variables that are likely to be used frequently by NCHS and other data users. They are edited for consistency, and missing values are imputed. Many (but not all) of these recoded variables are constructed from other variables in the NSFG; some are constructed from a large number of other variables. Other variables in the data file are not edited or imputed in this way.

Replicate—A probability subsample of the full sample design. The complete sample consists of several replicate subsamples, each of which is a small national sample of housing units. Replicate samples are released over the data collection to control the workflow of the interviewers. In responsive designs, early replicates are used to measure key cost and error features of a survey.

Respondent—A person selected into the sample who provides an interview. In the 2006–2010 NSFG, the ‘‘respondents’’ are the approximately 5,500 men and women 15–44 years of age who completed the NSFG interview each year.

Response rate—Respondents to a survey divided by the number of eligible persons in the sample. In this report, the response rate is the number of respondents (15–44 years of age) divided by the number of eligible persons (15–44 years of age). Given that not all screeners were completed, the number of eligible persons is not known precisely, so this number is estimated.

Responsive design—Responsive survey designs pre-identify a set of design features that could affect costs and errors of survey statistics; identify a set of indicators of the cost and error properties of those features; monitor those indicators in the initial phases of data collection; alter the active features of the survey in subsequent phases based on cost and error tradeoff decision rules; and combine data from the separate design phases into a single estimator.

Sample line—‘‘Sample line’’ is a ‘‘hold-over’’ term from an era in which interviewers were sent to selected area segments (blocks or linked groups of blocks) to list all housing units. The listing was done on paper, and later keyed to a master list. The sample for any given survey was selected from the master list. The housing units listed were ‘‘lines’’ on the listing sheet, and the terminology was applied to the electronic records in the master list.

The current design primarily uses U.S. Postal Service DSF addresses obtained from a commercial firm in each segment. In segments where the commercial firm cannot provide adequate numbers of addresses (for example, in rural areas where rural delivery routes are used and no house numbers or street names are available in the DSF), listing is done ‘‘from scratch.’’ Interviewers visit these segments and list all housing units directly into a laptop. Listed addresses are uploaded to the central office at the end of each day of listing. The ‘‘master file’’ contains addresses from the DSF and from scratch listings. On occasion the term ‘‘sample lines’’ is used to refer to the electronic records in this file. Thus, sample lines are addresses and not necessarily housing units. They become sample housing units once selected and households when the interviewer visits and finds the housing unit occupied.

Sampling variance—The sampling variance is a measure of the variation of a statistic, such as a proportion or a mean, which is due to having taken a random sample instead of collecting data from every person in the full population. It measures the variation of the estimated proportion or mean over repeated samples. The sampling variance is zero when the full population is observed, as in a census. For the NSFG, the sampling variance estimate is a function of the sampling design and the population parameter being estimated (for example, a proportion or a mean). Many common statistical software packages compute ‘‘population’’ variances by default; these may underestimate the sampling variance. Estimating the sampling variance requires special software, such as those discussed in this report.

Sampling weight—For a respondent in the NSFG, the ‘‘sampling weight’’ is the estimated number of persons in the target population that he or she represents. For example, if a man in the sample represents 12,000 men in his age and race and ethnicity category, then his ‘‘sampling weight’’ is 12,000. The NSFG sampling weights adjust for different sampling rates (of the age and race and ethnicity groups), different response rates, and different coverage rates among persons in the sample, so that accurate national estimates can be made from the sample. Because it adjusts for all these factors, it is sometimes called a ’’fully adjusted’’ sampling weight.

Screening interview—Sometimes called a ‘‘screener,’’ a screening interview is a (usually short) set of questions asked of a household member to determine whether the household contains anyone eligible for the survey. In the NSFG, the screening interview consisted of a household roster, collecting age, race, ethnicity, and gender. Those households having one or more persons 15–44 years of age were eligible for a main interview. In the NSFG, only persons 18 and over can answer the screener.

Self-representing area—A county or group of counties forming a primary sampling unit with population counts sufficiently large to be equal to or greater than the typical stratum size in the U.S. national sample. Such PSUs are thus represented in all draws of a national sample using the design. The sampling probabilities for persons in such areas are designed to be equal to that applicable in smaller PSUs, called nonself-representing areas.

Segment—A group of housing units located near one another, all of which were selected into the sample.

Simple random sample—A sample in which all members of the population are selected directly and have an equal chance to be selected for the sample. The NSFG sample is not a simple random sample. The NSFG sample was stratified, selected in stages, and employed unequal chances of selection for the respondents, by age, race, ethnicity, and gender. Such designs are referred to as ‘‘complex’’ and require special software to estimate the variance of statistics computed from a sample with a complex design.

Strata; Stratification—Stratification is the partitioning of a population of sampling units into mutually exclusive categories (strata). Typically, stratification is used to increase the precision of survey estimates for subpopulations important to the survey’s objectives. In the 2006–2010 NSFG, those groups include teenagers (15–19 years of age), Hispanic men and women, and non-Hispanic black men and women. To obtain larger and more reliable samples of these groups, the

NSFG sample was stratified: in the first stage of selection, PSUs were stratified using socioeconomic and demographic variables; in the second stage of selection, segments within each PSU were stratified by the proportion of the population that was black or Hispanic.

SurveyTrak—A software-based sample administration system that was used in the 2002 and 2006–2010 NSFG. The system is used by interviewers on laptop computers to document their sample assignment, to organize the activities of their workday, to prompt them for appointments to be kept, to record results of each call attempt, to record observations of the sample housing unit, and otherwise keep track of their job duties.

Target population—The population to be described by estimates from the survey. In NSFG the target population was the household population of the United States, which refers to the civilian noninstitutionalized population, plus active-duty military who are not living on military bases. ''Noninstitutionalized'' refers to the omission of prisons, hospitals, dormitories, and other large residences under central control. College students living in dormitories were interviewed but sampled through their parent or guardians' households.

UM-ISR—the University of Michigan Institute for Social Research.

WEBDOC—A software-based presentation of metadata and other survey documentation used for the NSFG. The NSFG Webdoc shows the variable names, category labels, and frequencies for each variable in the data file. The NSFG Webdoc is accessible at: http://www.cdc.gov/nchs/nsfg.htm.

Weight—See ''Sampling Weight.''

Appendix II. Degrees of Freedom in Estimates from the National Survey of Family Growth

As noted earlier in the report, the NSFG design is based on a national sample containing 110 primary sampling units (PSUs). For purposes of variance estimation, these 110 PSUs are rearranged to create sampling error computing units (SECUs) that parallel the sample design, mask the identity of geographic areas representing the PSUs to protect respondent confidentiality, and produce variance estimates for the stratified multistage design that are quite close to those that would be obtained if the 110 PSUs were used directly. These SECUs will be more numerous than the PSUs, as noted in the estimation section of this report.

For the expected NSFG 2006–2010 release, all 110 PSUs will be included. This release thus includes eight ''super 8,'' 20 other self-representing (SR), and 82 nonself-representing (NSR) PSUs. These are assembled into a set of 152 SECUs: 32 from the super 8, 40 from the SR PSUs, and 80 from the NSR PSUs, or a total of 152 SECUs. (4 of the 82 NSR PSUs will be combined to form two SECUs, thus creating 80 SECUs from 82 PSUs.)

These SECUs are grouped into 56 sampling error strata: 16 for the ''super 8'' PSUs containing two SECUs each; 20 for the SR PSUs containing two SECUs each; and 20 for the NSR PSUs containing four SECUs each. Thus, there are 16 + 20 pairs of units for the super 8 and SR part of the sample, and 20 ''quads'' for the NSR portion of the sample.

The NSR grouping into strata with four SECUs is done because of the need to provide calendar year estimates. Any one calendar year spans two national samples, and thus uses 20 NSR PSUs from each calendar year. Pairs could be formed for any two successive data collection years to group the 40 total NSR PSUs into strata for variance estimation purposes. However, a NSR SECU in one data collection year would have to be paired with an NSR SECU in the year before and the year following. This would require multiple sets of variance estimation strata and SECUs that users would then have to select on the basis of what data collection and calendar years are being used in a particular analysis. This is too complicated a scheme to impose on NSFG users. Instead, the ''quad'' grouping of strata provides a nonstandard but statistically sound way to provide a single set of sampling error strata and SECUs for any release or calendar year estimate.

The ''quad'' grouping does incur a slight loss in precision. ''Quad'' grouping requires a higher level of collapsed strata than ''pairs.'' For example, there are 80 total SECUs in 80 total strata in the sample of NSR PSUs. In a pairwise collapsing, the 80 stratum would be collapsed into 40 and 40 stratum boundaries ignored. This collapsing leads to a slight over-estimation of variance. On the other hand, 20 ''quads'' ignore an additional 20 stratum boundaries over the pairwise collapsing by collapsing four strata into one sampling error stratum. By ignoring the additional stratum boundaries, there is a further increase in estimated variance. This additional increase has been examined empirically, and preliminary analysis suggests that the increase in variance is in the range of 7 to 8 percent for means and proportions, or 3 to 4 percent increases in estimated standard errors. Such modest losses can be tolerated for many NSFG estimates. Therefore, for user convenience of a single set of strata to be used across all possible analyses, the NSFG uses a grouping of four SECUs in the NSR portion of the sample.

The degrees of freedom for an analysis depend on the number of SECUs and the number of strata used in the analysis. The full 4-year sample in the 2006–2010 release will contain 152 SECUs created from the 110 PSUs grouped into 56 strata (16 from the super 8, 20 from the SR, and 20 ''quads'' from the NSR). The degrees of freedom can be computed approximately as the number of SECUs minus the number of strata, or a total of 152 – 56 = 96 degrees of freedom. This degrees of freedom count will only be available for analyses of the complete 2006–2010 release. Shorter time periods, such as the 2-year 2007–2008 calendar year period will have fewer degrees of freedom.

For example, a single data collection year will have the full set of 32 variance computation units formed for the super 8 PSUs. There will be five rotating SR SECUs each data collection year generating 10 SECUs, and there will be 20 rotating NSR SECUs. Thus, a single data collection year from July in the first year through June in the second will have 62 SECUs. Since at present there are no plans to produce data collection year estimates, these units will not need to be grouped into sampling error strata, and degrees of freedom are not relevant.

However, in a single calendar year, which spans two data collection years, there will be 62 SECUs from the first data collection year plus 30 new SECUs from the following data collection year, a total of 92 units. Average counts per unit indicate cluster sizes for female or male data sets for a calendar year will be small (20 males and 25 females per unit) reducing design effects for calendar year estimates. With respect to degrees of freedom, the 32 super 8, 20 SR SECUs, and 40 NSR SECUs will be grouped into 46 strata. Thus, an annual calendar year estimate will have 92 – 46 = 46 degrees of freedom.

For a two calendar year estimate, there will 16 degrees of freedom from the super 8 PSUs, 15 degrees of freedom from the 15 SR SECUs in three data collection years spanning the two calendar year period, and 60 – 20 = 40 degrees of freedom from NSR SECUs, or a total of 71 degrees of freedom.

Finally, research continues concerning the reliability of estimates for the new variables added in 2007. In the 2006–2008 release, a weight will be provided for the 18-month period during which the new variables were collected. The degrees of freedom for this 18-month sample will be based on interviews in all eight super 8 PSUs, 10 SR PSUs, and 40 NSR PSUs. For 5 of the SR and 20 of the NSR SECUs in such an 18-month period, the number of interviews will be one-half as large as those from the other 5 SR and 20 NSR, because only 6 months of interviewing will be represented. Nonetheless, the degrees of freedom for data estimates from such an 18-month sample will be the same as those for a two calendar year estimate, or 46 degrees of freedom.

Vital and Health Statistics Series Descriptions

ACTIVE SERIES

Series 1. **Programs and Collection Procedures**—This type of report describes the data collection programs of the National Center for Health Statistics. Series 1 includes descriptions of the methods used to collect and process the data, definitions, and other material necessary for understanding the data.

Series 2. **Data Evaluation and Methods Research**—This type of report concerns statistical methods and includes analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Also included are experimental tests of new survey methods, comparisons of U.S. methodologies with those of other countries, and as of 2009, studies of cognition and survey measurement, and final reports of major committees concerning vital and health statistics measurement and methods.

Series 3. **Analytical and Epidemiological Studies**—This type of report presents analytical or interpretive studies based on vital and health statistics. As of 2009, Series 3 also includes studies based on surveys that are not part of continuing data systems of the National Center for Health Statistics and international vital and health statistics reports.

Series 10. **Data From the National Health Interview Survey**—This type of report contains statistics on illness; unintentional injuries; disability; use of hospital, medical, and other health services; and a wide range of special current health topics covering many aspects of health behaviors, health status, and health care utilization. Series 10 is based on data collected in this continuing national household interview survey.

Series 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**—In this type of report, data from direct examination, testing, and measurement on representative samples of the civilian noninstitutionalized population provide the basis for (1) medically defined total prevalence of specific diseases or conditions in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics, and (2) analyses of trends and relationships among various measurements and between survey periods.

Series 13. **Data From the National Health Care Survey**—This type of report contains statistics on health resources and the public's use of health care resources including ambulatory, hospital, and long-term care services based on data collected directly from health care providers and provider records.

Series 20. **Data on Mortality**—This type of report contains statistics on mortality that are not included in regular, annual, or monthly reports. Special analyses by cause of death, age, other demographic variables, and geographic and trend analyses are included.

Series 21. **Data on Natality, Marriage, and Divorce**—This type of report contains statistics on natality, marriage, and divorce that are not included in regular, annual, or monthly reports. Special analyses by health and demographic variables and geographic and trend analyses are included.

Series 23. **Data From the National Survey of Family Growth**—These reports contain statistics on factors that affect birth rates, including contraception and infertility; factors affecting the formation and dissolution of families, including cohabitation, marriage, divorce, and remarriage; and behavior related to the risk of HIV and other sexually transmitted diseases. These statistics are based on national surveys of women and men of childbearing age.

DISCONTINUED SERIES

Series 4. **Documents and Committee Reports**—These are final reports of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002. As of 2009, this type of report is included in Series 2 or another appropriate series, depending on the report topic.

Series 5. **International Vital and Health Statistics Reports**—This type of report compares U.S. vital and health statistics with those of other countries or presents other international data of relevance to the health statistics system of the United States. The last Series 5 report was published in 2003. As of 2009, this type of report is included in Series 3 or another series, depending on the report topic.

Series 6. **Cognition and Survey Measurement**—This type of report uses methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999. As of 2009, this type of report is included in Series 2.

Series 12. **Data From the Institutionalized Population Surveys**—The last Series 12 report was published in 1974. Reports from these surveys are included in Series 13.

Series 14. **Data on Health Resources: Manpower and Facilities**—The last Series 14 report was published in 1989. Reports on health resources are included in Series 13.

Series 15. **Data From Special Surveys**—This type of report contains statistics on health and health-related topics collected in special surveys that are not part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002. As of 2009, reports based on these surveys are included in Series 3.

Series 16. **Compilations of Advance Data From Vital and Health Statistics**—The last Series 16 report was published in 1996. All reports are available online, and so compilations of Advance Data reports are no longer needed.

Series 22. **Data From the National Mortality and Natality Surveys**—The last Series 22 report was published in 1973. Reports from these sample surveys, based on vital records, are published in Series 20 or 21.

Series 24. **Compilations of Data on Natality, Mortality, Marriage, and Divorce**—The last Series 24 report was published in 1996. All reports are available online, and so compilations of reports are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 5412
Hyattsville, MD 20782

1–800–232–4636
E-mail: cdcinfo@cdc.gov
Internet: http://www.cdc.gov/nchs

**U.S. DEPARTMENT OF
HEALTH & HUMAN SERVICES**

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782

CS209449
T36729 (6/2010)
DHHS Publication No. (PHS) 2010–1350, Series 2, No. 150