

# **The National Survey of Family Growth**

**Report to the NCHS**

**Board of Scientific Counselors**

**and its NSFG Review Panel**

**September 14, 2009**



# TABLE OF CONTENTS

## **INTRODUCTION**

Note

The Purpose of the NSFG

The Transition to Continuous Interviewing

The NSFG team

## **SURVEY PLANNING AND QUESTIONNAIRE CONTENT**

Co-funding agencies

Obtaining outside advice

High-volume NSFG data users

## **DATA PRODUCTION AND SYSTEMS**

The survey and data production process in the continuous NSFG

Imputation

Compliance with federal mandates for information security

## **DISSEMINATION OF MICRODATA**

Public use data files for Cycles 4-6

Restricted-use data files

Integrated Fertility survey series

Public use data files from continuous interviewing

## **PUBLICATIONS**

Key Statistics from the NSFG

Publications from Cycles 1-6

Publications planned from Continuous Interviewing

## **OTHER USER SERVICES**

## **THE BUDGET**

## **FUTURE NEEDS AND DECISIONS**

## **APPENDIXES**

1. Series 1, No. 48: Planning and Development of the Continuous NSFG (extract; forthcoming in September, 2009)
2. Series 2: Continuous NSFG: New Concepts for Sample Design and Analysis (extract; forthcoming about January, 2010)
3. Publications from the 2002 NSFG by NSFG team staff
4. Known Publications from the 2002 NSFG, September, 2009

## INTRODUCTION

### Note

This report is a relatively brief overview of the National Survey of Family Growth, or NSFG. We recommend looking at the Vital and Health Statistics, Series 1 report (**Groves et al 2009**; an extract is included here as **Appendix 1**) for details such as questionnaire outlines, field procedures, further details on the NSFG's innovative sample design and survey management techniques, and a step-by-step account of the details of the data collection and informed consent process. That report is expected to be posted online in September. Further information on sample design, weighting, imputation and variance estimation is in the forthcoming Series 2 report (**Lepkowski, et al, 2009; Appendix 2**), which will be published later this year or early in 2010.

**Readers of this report are also encouraged to look at the survey web site** for further information, including:

- NSFG questionnaires and codebooks for all NSFG surveys since 1973;
- Lists of publications for all surveys;
- PDF files of NCHS reports based on the NSFG;
- Downloadable public use data files; and
- “Key Statistics from the NSFG.”

The NSFG web address is:

<http://www.cdc.gov/nchs/nsfg.htm>

The NSFG e-mail box is at [nsfg@cdc.gov](mailto:nsfg@cdc.gov)

### The Purpose of the NSFG

The National Survey of Family Growth (NSFG) responds to the requirement in Section 306 of the Public Health Service Act, that the National Center for Health Statistics (NCHS) “shall collect statistics on...family formation, growth, and dissolution,” in addition to “determinants of health,” and other aspects of health and health care. More substantively, the NSFG is the US national fertility survey. It has been housed in the Division of Vital Statistics (DVS) in NCHS since 1973 when the survey was launched as a federal activity. Surveys of married women were conducted by private-sector foundations in 1955 and 1960 and by university researchers with NIH funding in 1965 and 1970. The first NSFG conducted by NCHS was in 1973.

The NSFG has been located in DVS because its original and principal purpose is to explain trends and group differentials in birth and pregnancy rates -- by collecting data on sexual activity, marriage and cohabitation, contraceptive use, infertility, breastfeeding, intended and unintended births, and related topics. In this way, it directly supplements and complements the data collected from the birth registration system. It also provides nationally representative data on cohabitation, marriage, and divorce, serving as a partial replacement for the discontinued national marriage and divorce registration systems.

Since the NSFG began including never-married women in 1982 and an independent (unlinked) sample of men in 2002, the NSFG has collected more data on reproductive and sexual health, including data on STD/HIV risk behaviors, fatherhood involvement, and men's and women's attitudes toward marriage, parenthood, and gender roles.

In addition to providing data on factors related to birth and pregnancy rates for NCHS, the NSFG provides data on demographic and public health issues for the agencies that help NCHS support the survey. NCHS contributes just over one-fourth of the contract costs of the survey plus all of the staff and other personnel support.

The agencies supporting the NSFG are all from the US Department of Health and Human Services (DHHS). The major supporters of the survey (each of whom contribute about one-fourth of the NSFG's annual costs) are:

- NCHS
- The Center for Population Research of the NICHD/NIH; and
- The Office of Population Affairs.

These agencies have helped to fund and shape the survey since the early 1980's.

The other one-fourth of the survey's costs is contributed by:

- The Children's Bureau;
- The Assistant Secretary for Planning and Evaluation;
- The CDC's Division of HIV/AIDS Prevention;
- The CDC's Division of Sexually Transmitted Disease Prevention;
- The CDC's Division of Reproductive Health.

(Others, such as the CDC Office of Women's Health and the DHHS Office of Child Support Enforcement, made contributions in a single year, but are not current funders.)

As shown in Figure 1, the NSFG was conducted with national samples of women 15-44 years of age in 1973, 1976, 1982, 1988, and 1995. In 2002, an independent national sample of men was added. In each of these surveys, all interviewing was done in 6-12 months by hiring a large staff (often 200 or more interviewers) and completing the interviews as quickly as possible. NSFG interviews have always been conducted only in person, in respondents' homes. Interviews are conducted only by trained female interviewers, and may be conducted in either English or Spanish. Interviews averaged 60 minutes for males and 80 minutes for females, and included both interviewer-administered and self-administered components. The self-administered component was done using Audio Computer-Assisted Self-Interview (or ACASI). (**Groves et al, 2005; Martinez et al, 2006**). Here are a few salient facts about each cycle of the NSFG (Figure 1 and **Groves et al, 2009** show more details):

---

**FIGURE 1: NSFG Cycles, with Sample sizes**

<u>Year</u>	<u>Cycle</u>	<u>Sample</u>	<u>Contractor</u>
1973	1	9,797 women* 15-44	NORC
1976	2	8,611 women* 15-44	Westat
1982	3	7,969 women 15-44	Westat
1988	4	8,450 women 15-44	Westat
1995	5	10,847 women 15-44	RTI
2002	6	12,571: 4,928 men 7,643 women	Univ of Michigan Institute for Social Research

\*The 1973 and 1976 surveys excluded never married women unless they had children living with them as shown on the household roster. Never-married women without children were first included in the survey in 1982.

---

### **The Transition to Continuous Interviewing**

In 2000-2004, the NSFG team and our co-funding agencies discussed the roles that the NSFG served, and concurred that the content and design of the survey was now meeting our needs fairly well. However, the survey needed to be conducted and released more frequently, with larger sample sizes, at approximately the same annual cost. To do this, we would need to limit the amount of questionnaire change across survey years and devise a more efficient sample design and fieldwork organization to cope with an increasingly difficult survey environment, in which it is more difficult to find people at home, and more difficult to obtain cooperation. (Groves et al, 2009)

For the next 3 years (2001-2003), we met periodically to discuss a sample design proposed by our then-Project Director, Robert Groves of the University of Michigan. After extensive study and discussion of 3 possible designs, the agencies agreed, in March of 2004, to proceed with a continuous interviewing design for the NSFG. (The 3 designs were the Cycle 6 design, a 3-year semi-continuous design, and continuous interviewing.) After intensive staff and contractor efforts to make necessary questionnaire modifications and obtain IRB and OMB clearances as quickly as possible, continuous interviewing began about July 1, 2006. (More details on the planning process are given in Groves et al., 2009).

The new design is described in more detail later in this document, but it is working extremely well thus far. As of September 2009, more than 17,700 interviews have been completed since July 1, 2006 (9,600 women and 8,100 men). Of these, 1,290 interviews have been conducted in Spanish, and the rest in English. The weighted response rate is about 75%. Efficiency (measured in terms of “hours of interviewer labor per completed interview” or “HPI”) is 20% better than in 2002. In sum, it is fair to say that the early years of continuous interviewing in the NSFG have been a great success.

This success has been recognized by OMB and other federal agencies, as a model for other federal surveys. We have been asked to give presentations for the NCHS-Statistics Canada “Interchange Meetings in Nov 2005, Nov 2007, and Nov 2008; at the American Statistical Association (August 2009), at inter-agency forums sponsored by OMB (Nov 2008 and June, 2009), and at the Population Association of America in 2009, on the ways in which we are using paradata (data on the data collection process), a “double sample” for non-response follow-up, and incentives to increase the cost-effectiveness of the data collection and to manage costs and prevent cost increases. The NSFG is, then, an innovator in the management of survey fieldwork. (See also **Groves et al, 2005, 2009; Groves and Heeringa, 2006**). Further details can be provided on request.

**The NSFG Team**

All aspects of the NSFG, from survey design and fieldwork through production of data files and documentation, are conducted by the NCHS/NSFG team of 6 staff, working closely with a contractor (currently, the University of Michigan’s Institute for Social Research). The NCHS/NSFG team has varied from 4-8 over the past 30 years, but its current size (6) is typical.

The University of Michigan’s Project Director for the NSFG contract was, until April 2009, Robert Groves, now Director of the US Census Bureau; William Axinn is the Deputy Director, James Lepkowski, the senior Mathematical Statistician, and Nicole Kirgis, the Field Director and operations manager. A successor to Dr. Groves will be announced shortly.

The current members of the NSFG team are listed here in alphabetical order:

<u>NAME</u>	<u>AT NSFG SINCE</u>	<u>GRADE</u>
Joyce Abma, Ph.D.	1993	14
Anjani Chandra, PhD.	1991	14
Casey Copen, PhD.	2008	12
Jo Jones, PhD.	2004	13
Gladys Martinez, Ph.D.	2000	13
William Mosher, PhD.	1978	14

Joyce Abma’s roles include oversight of the design and implementation of the sections of the questionnaires on contraceptive use and intended and unintended fertility; serving as Alternate Project Officer for the contract; coordinating tasks related to producing the documentation for the public use files; and stewardship of older data files on the CDC mainframe. Her publications have dealt with sexual activity and contraceptive use among teenagers, estimates of pregnancy rates, voluntary childlessness, non-voluntary sexual encounters, and other topics.

Anjani Chandra’s roles include oversight of the design and implementation of the ACASI section of the questionnaires, and the sections dealing with pregnancy and male and female fertility histories, fecundity and infertility, and marriage and cohabitation. She also serves as coordinator of the specification and testing of all sections of both the male and female NSFG questionnaires

for programming, and for the production and testing of the recoded (“constructed”) variables for the public use files (including imputation of missing data). She makes important contributions to the preparation of NSFG materials for the NCHS Disclosure Review Board, OMB, and the IRB. Her publications have dealt with infertility and infertility services, sexual behavior and HIV risk, maternal and infant health, and other topics. In addition to her NSFG duties, Dr. Chandra serves as Vice Chair of the NCHS Research Ethics Review Board.

Casey Copen, our most recent staff addition, came to the NSFG in 2008. She is assuming increasing responsibility for questionnaire sections dealing with marital and cohabitation histories and social and demographic characteristics. Her research interests include teenagers, sexual behavior, reproductive health, and marriage and cohabitation histories.

Jo Jones was formerly project manager of the Add Health survey, and then worked for Abt Associates on the National Immunization Survey and various State and Local Integrated Telephone Surveys (SLAITS) projects before coming to NCHS in 2004. She oversees the design and implementation of sections of the questionnaires dealing with adoption, fatherhood, and such socio-demographic characteristics as income and religion. She also works with users of the NSFG’s restricted-use data files in the NCHS Research Data Center. She serves as NSFG liaison with the NCHS web site staff to request and review updates and upgrades in the NSFG web site. Her publications have dealt with adoption, sexual behavior, and fertility. She also led the planning for the 2006 and 2008 NSFG Research Conferences.

Gladys Martinez has oversight of the design and implementation of the sections of the questionnaires covering female family planning and birth control services, future birth expectations, male reproductive health care, and father involvement. She coordinates the translation, implementation, and testing of the Spanish-language Questionnaires and all related materials. She also serves as the NSFG’s principal liaison to the Healthy People Family Planning objectives work group. Her publications have dealt with teenage sexual activity and fertility, teen fatherhood, childlessness, and family planning, among other topics.

Bill Mosher is responsible for technical (substantive) oversight of the contract (this role is called a “Project Officer”; Joyce Abma is the “Alternate Project Officer”), and monitors the survey’s budget and funding situation. He also reviews contract vouchers, and often represents the survey to NCHS, the CDC, and the rest of the bureaucracy. He coordinates the team’s efforts in writing the NSFG contract, contract modifications, OMB clearance materials, IRB protocols, DRB review, and computer security documents. He coordinates the team’s efforts in planning the NSFG’s publication program, and works with funding agencies and outside experts on advisory meetings. His publications have dealt with most of the NSFG’s topics; his primary interests are fertility, contraception, and survey methodology.

In addition to these 6, the Reproductive Statistics’ Branch Chief, Stephanie Ventura, assists the NSFG team with review of publications, procuring staff and other resources, and interacting with higher level managers on behalf of the survey. The Director of DVS, Charles Rothwell, does similar tasks at a higher level. Karen Davis of the NCHS Office of Research and Methodology (ORM) also contributes about 10-20% of her time to provide advice on sample design, weighting, variance estimation and other statistical issues to the NSFG team.

## SURVEY PLANNING AND QUESTIONNAIRE CONTENT

### Co-Funding Agencies

The NSFG's content is arrived at in discussions with our collaborating (co-funding) agencies and selected outside experts. Because about ¾ of the funding from the NSFG comes from these outside sources, we make certain that we are meeting their data needs to the extent possible. Therefore we meet with representatives of those agencies at least once a year, and send them frequent progress reports.

Since Cycle 1 in 1973, the core of the NSFG's content for women has included: a birth and pregnancy history, a complete marriage history, and partial histories of cohabitation and contraceptive use (including month by month for the last 3-4 years). Since the 1980's, that content has also included data on nonmarital, noncohabiting sexual partners during the 12 months before the interview.

For men, who were first interviewed in 2002, a history of births fathered, and information on marriage and cohabitation, are collected, along with data on what the father does to help raise his children, whether they live with him or not.

For both sexes, data are collected on social and demographic characteristics, health conditions related to reproduction, attitudes toward marriage and parenthood, and (in Audio CASI) sexual and drug use behaviors that affect the risk of HIV and other sexually transmitted diseases.

In the Continuous NSFG, begun in 2006 and now underway, NCHS is collecting data to carry out its own responsibilities, and to assist our co-funding agencies in DHHS to carry out their own duties:

- the Office of Family Planning, Office of Population Affairs (OPA), under Section 1009 of Title X of the Public Health Service Act;
- the Adolescent Family Life Program of the Office of Adolescent Pregnancy Programs, Office of Population Affairs (OPA), under Section 2001 of the Public Health Service Act;
- the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Public Health Service, under Section 301 of the Public Health Service Act;
- the Division of HIV/AIDS Prevention (DHAP) of the National Center for HIV, Sexually Transmitted Disease, and Tuberculosis Prevention (NCHSTP, Section 301 of the Public Health Service Act);
- the Division of Reproductive Health (DRH) of the CDC, under Section 301 of the Public Health Service Act;

- the Division of Sexually Transmitted Disease Prevention of the Centers for Disease Control and Prevention (CDC), under Section 301 of the Public Health Service Act;
- the Children’s Bureau of the Administration for Children, Youth, and Families, Office of Human Development Services, under PL 96-272, the Adoption Assistance and Child Welfare Act of 1980; and
- the Office of the Assistant Secretary for Planning and Evaluation (OASPE), under Section 301 of the Public Health Service Act.

Across agencies, the NSFG is an important source of data for the “Healthy People 2010” and “Healthy People 2020” Objectives. For example, the NSFG is the data source for nearly all of the Healthy People objectives on Family Planning (13 objectives with numerous sub-objectives.) The NSFG is also the source of data for some objectives in the Healthy People focus areas on sexually transmitted diseases, HIV, and maternal and infant health.

### **Obtaining outside advice**

In addition to obtaining advice and suggestions from our collaborating (funding) agencies, and from researchers at professional meetings, we have also sought to obtain input on the content of the survey from a variety of outside experts.

We had annual meetings with an informal, rotating group of outside advisors between 1990 and 2002; many of these meetings were focused on questionnaire content because we made major changes in the questionnaires in Cycles 5 and 6. A brief account of the issues discussed at these meetings is given in **Kelly et al (1997)** and in **Groves et al (2005)**.

In November of 2008, we had another meeting with an informal group of outside experts. As in 1990-2002, these experts were chosen by NCHS, NICHD, and OPA (the 3 major funders of the survey). Representatives of each funding agency attended. The invited experts were:

1. Laura Lindberg, Ph.D., Senior Research Associate, the Guttmacher Institute, on behalf of Jennifer Frost, Ph.D. (both sociologist/demographers)
2. S. Philip Morgan, Ph.D., Professor of Sociology, Duke University (former President of the PAA; sociologist/demographer)
3. Elizabeth Peters, Ph.D., Department of Policy Analysis & Management; Director, Cornell Population Program, Cornell University (economist/demographer)
4. John Santelli, MD, MPH, Chair, Dept of Population & Family Health, Columbia University School of Public Health (pediatrician, epidemiologist)
5. Elizabeth Thomson, Ph.D., Dept of Sociology, University of Wisconsin, & Professor of Demography, Stockholm University (sociologist/demographer)

(Debra Kalmuss, Ph.D., Columbia University School of Public Health, was invited but unable to attend at the last minute because of a family emergency.)

At the November 2008 advisory meeting, the NSFG staff and Dr. Robert Groves, our then-project director, briefed the experts on the status of our fieldwork and the techniques we were using to manage fieldwork. The NSFG team then sought advice on:

- (1) questionnaire changes for the next survey, starting in 2011 in several specific areas, including sex education, fatherhood involvement, wantedness of births, attitudes, and other topics. Many suggestions were made. As other contract tasks allow, we are attempting to incorporate these into the questionnaire for 2011.
- (2) possible contract options, which had been in the Cycle 6 contract. None were pursued in Cycle 6 because of a lack of funding, and because the co-funders agreed in 2000-2004 that using continuous interviewing to produce larger samples and more frequent data releases was more important.
- (3) ways to improve our outreach to current and potential data users. We obtained a number of useful suggestions in this discussion, and we are implementing as many as we can for the 2006-08 public use file release.

In the Spring of 2009, the NSFG staff had a separate meeting with Drs. Kalmuss and Frost (who had been unable to attend the November meeting), and Arik Marcell, MD, a physician at the Johns Hopkins University who has experience working with the reproductive health care needs of male teens and young adults. This meeting was held to discuss issues and solicit suggestions related to the NSFG questionnaire sections that collect data on reproductive health services from men and women.

### **High-volume NSFG Data Users**

A special category of outside advisors consists of three organizations that have used NSFG data extensively over the past decade or more. The NSFG public use file has a diverse set of users, in sociology and economics departments, schools of public health, non-profit research organizations, and many state and federal agencies. But these organizations each have multiple researchers using the NSFG for analyses that are published and distributed to a wide variety of users. We therefore pay special attention to feedback from them about the usefulness of our data. They are:

#### **1. The National Campaign to Prevent Teen and Unplanned Pregnancy**

([www.thenationalcampaign.org](http://www.thenationalcampaign.org)) The Campaign was founded in 1996 and works with state and local governments and private groups on programs to prevent teen pregnancy, and more recently, on programs to prevent unplanned pregnancy among unmarried adults in their 20's. The Campaign makes extensive use of NSFG data, as can be seen from the Cycle 6 list of publications (**Appendix 4**), which lists at least 15 National Campaign publications that have used NSFG data. In 2006, the Campaign staff said that their publications and powerpoint

presentations that used NSFG had been downloaded (or distributed in hard copy) a total of 880,000 times, and used to some extent in most of the presentations and policy briefings they give. In recent years, the Campaign has expanded their mission to include unintended pregnancy to unmarried women in their 20's, and their use of our data has expanded accordingly.

**2. The Guttmacher Institute** (formerly the Alan Guttmacher Institute), New York.

This non-profit research organization has obtained a number of grants from NICHD, from the Office of Population Affairs (OPA), and from other sources to analyze NSFG data. Since the 1970's, they have published many reports, articles in professional journals, Fact Sheets, and other materials using NSFG data. In recent years, their web site ([www.guttmacher.org](http://www.guttmacher.org)) also makes extensive use of NSFG data. For two prominent examples of their research using the NSFG, see Finer and Henshaw (**reference 18**) and Kost et al (**reference 19**).

**3. Child Trends** of Washington, DC is a non-profit, non-partisan research organization ([www.childtrends.org](http://www.childtrends.org)) that does contract and grant research for government agencies (including OPA, NICHD, and others) and private foundations, and maintains an extensive "data bank" ([www.childtrendsdatabank.org](http://www.childtrendsdatabank.org)) that includes data from a variety of sources, including the NSFG. Their research is, as their name implies, focused on teens and young adults. They have used NSFG data extensively since the 1980's, when we began interviewing unmarried teenagers.

## DATA PRODUCTION AND SYSTEMS

In 1993 with the Cycle 5 Pretest, the NSFG became the first NCHS survey to begin collecting data using computer-assisted personal interviewing (CAPI) and Audio Computer-Assisted Self-Interviewing (ACASI). The CAPI and ACASI programs were revised and used again to interview 10,847 women in the Cycle 5 main study in 1995. Both surveys used an early version of BLAISE (a questionnaire programming language programmed by the Netherlands Central Bureau of Statistics; it is the most widely used CAPI programming language). The main study was administered in both English and Spanish.

Use of CAPI and ACASI continued in the 2002 (Cycle 6) survey, this time expanding to two different questionnaires—one for men and one for women. The survey took advantage of advancements in hardware and software. Interviewers used lightweight tablet computers with touchscreen capability for electronic storage of informed consent signatures. The questionnaire was programmed using advanced versions of BLAISE, allowing for a smoother interview with a lower rate of inconsistent data. Both the male and female questionnaires were administered in both English and Spanish.

Revising, specifying, testing, and implementing these questionnaires were formidable tasks for a survey with a staff of 6, along with about the same number at our contractor. The contract costs, NCHS staff time, and calendar time involved in specifying and testing these questionnaires led us and our co-funding agencies to search for more efficient ways of keeping the questionnaire current and responsive, by using a more stable core instrument and making smaller changes every year or two. This is the strategy we have followed in the continuous NSFG, with changes

made in 2006 (as continuous interviewing began) and much more limited changes made in 2007 and 2008.

### **The survey and data production process in the continuous NSFG**

In the continuous NSFG, the interviews are conducted by about 40 interviewers in any given year, in 33-35 areas (Primary Sampling Units or PSU's) around the country. **(Groves et al, 2009)** As each interview is completed, the data are immediately encrypted, and sent electronically, every night, to the contractor's headquarters, where they are stored on a separate, dedicated NSFG server. Every month, we and our contractor work together to perform consistency checks and other quality control tasks. We also begin the programming and checking of previously specified recoded (constructed) variables (e.g., Current Contraceptive Status, or Fecundity Status.).

Beginning in Cycle 6, we and our contractor developed a system to create the codebook for the datafile, using Excel spreadsheets, which we continue to use. The Excel spreadsheets are initially filled by automated output of variables from the BLAISE questionnaire. Working from the complete list of items output, separate spreadsheets for different components of the codebook documentation are created. These components include short descriptions or "short labels" for indexes of file content; the question wording (i.e., "question text"); a code indicating the public versus in-house status of the variable; and a specification of which respondents were asked the question (i.e., "applicable specifications" or "app specs").

These spreadsheets, along with a set of value labels for each variable (i.e., labels describing the response categories), are combined and merged with the data to create the codebook. Preliminary versions of the codebook are reviewed by NSFG staff and changes and corrections are made, resulting in the final codebook after as many revisions as needed. The codebook process for Continuous Interviewing began with the spreadsheets that were developed for Cycle 6. (The continuous interviewing design makes maximum use of previously existing materials, so only updates will be needed for subsequent data releases.)

When the codebook is ready, the NSFG team submits the documentation to the NCHS Disclosure Review Board or DRB. The DRB reviews the codebook and a detailed memo called a "Disclosure Review Checklist," and recommends changes in the file to reduce the risk of disclosure. Those changes are made to the data and documentation as needed. A User's Guide is written to accompany the codebook. This Guide includes all the information needed to use the data file.

The sampling weights (the number of people in the population that each sample person represents) are produced by the contractor and checked by the NSFG team and the ORM statistician. Variance estimation procedures and examples are written up and prepared for inclusion in the Series 2 report and the NSFG web site. **(Lepkowski et al, 2006, 2009)**

Since Cycle 3 (the 1982 survey), the highest priority has been placed on preparing a public use data file as soon as possible. One report on a high-profile topic (usually contraceptive use, or teen sexual activity) is released with the data file, so that users can check their results against

those in the published report. As soon as the data file and documentation are ready and have passed confidentiality review, the data file is uploaded to the NCHS web site and copied onto CD-ROM's for distribution.

## **Imputation**

The recoded variables (“recodes”) described earlier are the variables NCHS uses to make our national estimates of the most important outcome variables (contraceptive use, infertility, age at first intercourse, etc). To produce consistent national estimates, missing data for these variables (and only these variables) are imputed.

In Cycle 1, missing data for all variables were imputed in the NSFG, but it took too long, and too much staff time, to get the public use file ready. In Cycle 2, no variables were imputed, and NCHS and other users had persistent problems with inconsistent estimates of key variables. In Cycle 3, we settled on a middle ground: the recoded variables, and only the recodes, would be imputed. This can be done largely by the contractor, with minimal impact on the schedule for the public use file release, and still maintains consistent estimates of the most important outcome variables for official NCHS reports. It also helps outside users replicate NCHS results.

The level of missing data in most NSFG recoded variables is rather low -- about 1 percent or less for most variables and less than 2 percent for nearly all. (Income and a few others are exceptions) The NSFG uses a regression imputation approach for imputation, with logical imputation based on other questionnaire data for some cases. The approach is described in the Series 2 reports. (Lepkowski et al, 2006, 2009)

## **Compliance with federal mandates for information security**

The NSFG has a smaller-scale computer system than the NHIS, the NHANES, or the National Vital Statistics System. However, that does not lessen the stringency of the requirements to protect the security of the data or the need to document how we do so. Thus, complying with the Federal Information Security Management Act (FISMA) is especially challenging for a staff the size of the NSFG. Because all CDC systems have to be reviewed at once, this review process has overloaded the CDC staff brought in to do this work, so progress has been slow.

The NSFG's effort began in 2006 and is ongoing. First, NSFG staff attended training and filled out the initial documents related to Information security. The University of Michigan staff took on this task for the later documents. We have been working through the “Certification and Accreditation” process for IT (Information Technology) systems for the last 3 years. We are told that Certification and Accreditation (“C&A”) approval is imminent, as soon as the CDC Office of the Chief Information Security Officer (OCISO) provides feedback on some of our proposed solutions.

## DISSEMINATION OF MICRODATA

### **Public use data files from previous Cycles**

Preparing public use files that are clear, well-documented, and easy to use, and releasing them as soon as possible after the end of data collection, has been a high priority for the NSFG team since the 1980's. Given the many interrelated histories in the NSFG interview (birth and pregnancy history; marriage history; cohabitation history; contraception history) this work is challenging and time-consuming.

Given the size of the NSFG staff and the complexity of the data files, it took 19 months in Cycle 3 to get the data file ready for release; in Cycle 4, it took 18 months; and in Cycle 5, it took 18 months (although Cycle 5 was a much larger file than Cycle 4). In Cycle 6 (2002), it took about 21 months to release the public use data file, because of the need to process and document different data files for two data files—one for females and a significantly different file for males—with several thousand new variables.

In continuous interviewing, we expect significant improvements in timeliness. We expect that the data file containing interviews for July 2006-December 2008 will be released about 12-14 months after data collection ends (January 2010 would be 13 months). We expect that the file containing interviews through June of 2010 may be released even faster—about 10 months.

This is all being done without compromising the NSFG's high standards of documentation, which include:

- documenting the variable name, value label, question text, and applicable specifications for every item,
- providing complete specifications for all recodes,
- imputation flags for all variables with any imputed cases,
- user-friendly, searchable web-based documentation (called "Web-doc") as well as a paper-style PDF codebook;
- a detailed User's Guide, and
- editing and recoding as required by the NCHS Disclosure Review Board.

The NSFG staff makes an unusual effort to document the data files very thoroughly. Many of our data users tell us that they appreciate the efforts we make to make our data files easier to use.

### **Restricted-use data files**

The NSFG is a national fertility survey, designed to produce national estimates by age, sex, race, Hispanic origin, and other characteristics. Researchers seeking to explain contraceptive use, fertility, marriage, and other behaviors with geographic variables--such as state policies or the socio-economic characteristics of neighborhoods--have several options for using the NSFG, and these options are improving over time.

In both Cycle 5 and Cycle 6 (1995 and 2002), the NSFG contractor, in consultation with the NSFG staff, prepared a contextual data file using Census data and other data that can be linked

with the NSFG public use file. In fact, the need to house the NSFG contextual data file was one of the reasons the NCHS Research Data Center was created in the 1990's.

The NCHS Disclosure Review Board has decided, in both Cycle 6 and in continuous interviewing, to remove any geographic identifiers (except a metropolitan-non-metropolitan variable) from the NSFG public use files to reduce the risk of disclosure that it sees as stemming from the NSFG's extensive birth histories and marriage histories, combined with other socio-demographic characteristics.

Therefore, any user wanting the state of residence of NSFG respondents, or characteristics of the county or census tract in which they live, must use restricted-use data files through the NCHS Research Data Centers in Hyattsville or in Atlanta, or in the Census Research Data Centers (so far, NSFG users have used the centers in Chicago, Ann Arbor, Michigan, and New York). Some of these users are economists who wish to add state-level policy variables to the data file. Others are sociologists and others who want to study the effects of neighborhood characteristics on behaviors such as teenage sexual activity, cohabitation, marriage, and divorce, and contraceptive use.

Other variables that cannot be included on public use files for disclosure reasons are also available through the Data Center. For example, the household rosters (i.e., listings of all persons who "usually live" in the household, their exact ages and relationships, etc) also had to be excluded from the public use files, and several variables were collapsed in some way to reduce the risk of disclosure, but their full detail is still available through the RDC.

### **Integrated Fertility Survey Series**

The six (soon to be seven) data sets from the NSFG have had a core focus, even as changes, additions, and enhancements were made over time to improve the utility and value of the survey. An important initiative is now ongoing to prepare a consolidated file of the NSFG and its predecessor surveys back to 1955. The University of Michigan (Pamela Smock, Principal Investigator) received an NICHD grant to prepare a harmonized data file of fertility survey data from 1955 through 2002. The NSFG staff has consulted with this team of researchers on its work. Some of these files will be released on the web in 2009 and 2010. The IFSS web site is at: <http://www.icpsr.umich.edu/IFSS/>

### **Public use data files from Continuous Interviewing**

As soon as possible, probably in January of 2010, we expect to release the first public use data file based on interviews conducted from July 1, 2006-December, 2008.

The sample size for that data file is expected to be 13,495, compared with 12,571 in Cycle 6 (collected in 2002 and released in 2004), as shown below:

	<u>2002</u>	<u>2006-8</u>	<u>Increase</u>
Total	12,571	13,495	+925
Male	4,928	6,139	+ 1,211
Female	7,643	7,356	-287
Teen	2,271	2,785	+ 514
Black	2,460	2,451	-9
Hispanic	2,712	2,874	+162

Note that the 2006-8 data file will have (by design) a larger male sample size and a larger sample of teenagers than in 2002.

Interviewing using the current sample design is expected to continue through June, 2010, at which time the entire 4-year sample design will have been used. If interviewing continues to go well, total sample size for 2006-2010 would be approximately 22,000 interviews—by far the largest NSFG ever done. Another public use data file will be released in 2011 to permit use of the combined data file with all 22,000 interviews. It will also be possible to use sub-sets of the data file; specific guidance for data users will be released with the data files.

(Note that, although it took longer, the cost of completing these 22,000 interviews will be approximately the same as the cost of Cycle 6, in which more of our funds were spent on fieldwork, questionnaire programming, and a large pretest/pilot study. In Cycle 5, again more of the funds were spent on questionnaire programming and testing, on a large pretest, and on the costs of working a very dispersed sample design in 198 PSU's).

The NSFG team's view is that the anticipated cost efficiencies of this design and fieldwork approach are being realized, and the result will be a significantly larger and more useful dataset.

## PUBLICATIONS

### **Key Statistics from the NSFG**

This is our newest data product. Housed on the NSFG web site at [www.cdc.gov/nchs/nsfg/abc\\_list.htm](http://www.cdc.gov/nchs/nsfg/abc_list.htm), "Key Statistics" is an alphabetical list of basic findings on key topics from the 2002 NSFG. About 80 entries are shown, from "Adoption and Nonbiological parenting" to "Wanted pregnancy." Each entry gives a basic statistic, along with the source of the statistic, and a link to the report in which it was published, so that the reader can find more details. These statistics are meant to be an easy way for data users (such as students, teachers, and reporters) to find the statistic they need and to find the back-up information in the original source if they want to explore the topic further.

### **Publications from Cycles 1-6**

The NSFG web site contains bibliographies, updated every few months, of NCHS publications, and journal articles found through literature searches, that use the NSFG data files. (It does not include newspaper or magazine articles that merely cite NSFG data published elsewhere.) The

searches that produced this list are being updated in September 2009. When we have updated results, we will share them with the BSC and post them on the NSFG web site.

Publications in those lists include:

Cycles 1-2 (1973 & 76):	126 reports and articles
Cycle 3 (1982)	117
Cycle 4 (1988)	99
Cycle 5 (1995)	171
Cycle 6 (2002)	149 (so far)
<b>TOTAL</b>	<b>662 (as of August 2009)</b>

### **Publications planned from Continuous interviewing**

The NSFG team's publication plans for the next data file are based on our experience in publishing NCHS reports and journal articles from previous cycles. In Cycles 3 and 4, we published a series of several short, highly focused Advance Data Reports.

For example, in Cycle 4 (the 1988 survey), we published Advance Data No. 182 (current contraceptive use), 184 (family planning services), 189 (unwanted births), 192 (fecundity and infertility), 194 (marital dissolution), and 200 (HIV-related knowledge), in a 14-month period from March 1990 to May 1991.

These reports often had just 4-6 tables and they were 12 printed pages or less (including appendixes). While these reports make good "news story units" because they have just a few principal findings, we found that they had at least 3 disadvantages:

- (a) they were a great deal of work to release just a few tables,
- (b) they prompted waves of data requests for other data, and these requests overwhelmed our small staff, and delayed our work on subsequent reports and other tasks.
- (c) Finally, they tended to convey the inaccurate impression that the NSFG had only a limited number of outcome variables of interest. Many readers (including researchers) mistakenly thought that the NSFG did not have the outcome variables they wanted to study.

To prevent that situation, in Cycle 5 we published one large Series 23 report in May 1997 (91 tables), followed in 1998 by a special issue of a peer-reviewed journal with 6 NSFG articles. That worked well to get some of the basic findings out relatively soon after the data were released, and to show the wide range of data on the NSFG.

But it did not do justice to a number of other important topics, such as trends in teen sex and contraception (that report was published in 2001), and marriage and divorce (it took until July 2002 to publish our report on the 1995 data on marriage, divorce and cohabitation. That report is

the 9<sup>th</sup> most downloaded NCHS report in Fiscal Year 2009, with 16,000 downloads, despite the fact that the data were collected in 1995.

For the 2002 (Cycle 6) NSFG, we were acutely aware that we would have greatly expanded content on men, HIV risk, sexual behavior, attitudes, and other topics, and we wanted to adequately represent the NSFG's diverse new content. So we pre-planned and published several reports that collectively contained more than 300 tables of data (**see references 8-16**). These served as very useful references for NCHS and our other funding agencies in answering data requests, and giving presentations. Also, the reports helped students, teachers, and researchers learn about the broad scope and many research possibilities of the data set.

It appears that the strategy worked: a recent "Google Scholar" search revealed that some of these reports (especially references 8, 10, and 11) are very often cited in other published articles. The disadvantages of this approach were that it was a tremendous strain for a staff of 6 to get such large reports out; and it took more than 3 years after the end of data collection to get some of them published.

In the next data release, beginning in January 2010, we are taking a middle ground between these approaches. We are also recognizing that it has been 5 years since we last released a data set (Cycle 6 was released in December 2004), but in the future, our data will be released more frequently than in the past (every 2-4 years instead of every 7 years). As continuous interviewing becomes routinized, we can develop a standard set of reports that would be issued with each data release. We see this first set of reports, however, as a transition between the periodic surveys of the past (done every 7 years) and the more frequent releases of the future (every 2-4 years).

First, we are publishing our methodology reports (**Groves et al, 2009; Lepkowski et al, 2009**) with or before the data file. This way, the information in these reports will be out when the data are released, so that researchers will have the information they need to plan their analyses. These methodology reports will also not compete for our time with the reports of findings, as they did in previous cycles.

Second, the substantive reports will be smaller and more focused than in Cycle 6, but larger and more comprehensive than in Cycles 3 and 4. They would still serve many of the same purposes:

- for users who just want "a quick statistic" such as journalists, policy-makers, and others;
- for those who want basic findings to illuminate a broad topic, such as teachers seeking lecture material or students;
- for health or social policy-makers looking for in-depth information on key topics; and
- for researchers looking for interesting new findings to investigate further.

The reports are expected to be released in the following order, which takes into account their "news value" and the degree of difficulty in producing them. If these reports are released beginning in January 2010 at the rate of one every 2 months, they would be released between January of 2010 and May of 2011. (At that time, we would have to begin working on reports from the combined 2006-2010 data file.)

1. Sexual Activity, Contraceptive Use, and Pregnancy among teenagers. (Similar to **Reference 9**, Series 23, No. 24; about 30 tables).
2. Use of Contraception and Family Planning Services in the US. (Similar to **Reference 8**, Advance Data No. 350; about 25 tables.)
3. Sexual Behavior and HIV Risk in the US. (similar to a combination of **References 10 and 14**, Advance Data 362 and 377; about 30 tables.)
4. Intended and Unintended Births in the US: Trends and New Measures. (Revision of a poster given by Abma, Jones, & Mosher at PAA in 2008; about 15 tables.)
5. Fertility and Parenthood. (selected tables on these topics, similar to some of the tables published in **References 11 and 12**; including female and male fertility, child care, and father involvement. About 30 tables.)
6. Pregnancy Rates in the US (Similar to **Reference 15** and our other previous reports in the NVSR series over the last 20 years, which report birth, abortion, and fetal loss rates by age, race, Hispanic origin, and marital status for the US. This will continue the time series begun in 1976. About 10 tables.)
7. If time is available: Marriage, Divorce, and Cohabitation. (About 30 tables. Using data from 1995, 2002, and 2006-8 for women, and 2002 and 2006-8 for men.) If time is not available, this report will be prepared with the 2006-10 data file. (Interest in this topic is high, as noted above; the Cycle 5 report was downloaded 16,000 times in FY 2009.)

## OTHER USER SERVICES

We make a number of other regular presentations to familiarize potential users with the findings and potential uses of the NSFG, and to provide information that will be useful to users of the data files. These presentations include:

- NSFG Workshops at PAA (virtually every year since the early 1990's)
- NSFG Poster presentations at American Sociological Association (in most years)
- NSFG workshops and presentations at the annual meeting of the MCH-Epi (Maternal and Child Health Epidemiology) conference (in most years)
- NSFG sessions at the NCHS Data User's Conference for many years
- NSFG sessions at the ICPSR Marriage & Family Data Workshop in 2008 and 2009
- Presentations at the OPA's annual "Title X Family Planning National Grantee" Meeting (most years, including 2007, 2008, 2009)

We also held NSFG Research Conferences at NCHS in 1998, 2006, and 2008. In each case, about 20 papers were presented and discussed by researchers from across the country. Presenters

included graduate students and faculty, and researchers from other organizations such as the Guttmacher Institute and Child Trends.

## THE BUDGET

The NSFG's contract budget has been about \$3.6 million per year in recent years. Of that, NCHS contributes about \$1 million and other agencies contribute the other \$2.6 million. That makes the NSFG smaller than the agency's largest surveys. The NSFG raises the other \$2.6 million a year from 7 other agencies, including NICHD, OPA, ASPE, Children's Bureau, and CDC's DRH, DHAP, and DSTDP. Funding so far appears adequate to finish interviewing under the current sample design in June of 2010, if all agencies contribute their promised amounts.

## FUTURE NEEDS AND DECISIONS

The NSFG will soon be going into a process to select a contractor for the next several years of the continuous NSFG. That process creates constraints on what can be said at a public meeting like the BSC. This is what we can say. (This BSC report will be included as an attachment to the RFP so that all offerors will all have the same information.)

A more secure funding source would allow NCHS to carry out the NSFG on schedule, and expand the sample size--either in the current age range (15-44 years of age), or to expand the age range to 15-49 or 15-54. An expansion of the age range would make our data on marriage and divorce, cohabitation, infertility and infertility services, father involvement, adoption and step-parenting much more useful. For example, expanding the age range to 54 would allow a much better answer to questions such as these, for the total population and for important demographic and socio-economic sub-groups:

“How many marriages will end in divorce?” or

“What proportion will remarry within 10 years after their first marriage ends?”

“What proportion of men and women ever adopt children, or ever raise a stepchild?”

Other options, such as a military or prison sample, a longitudinal follow-up, and others, are scientifically interesting and could have program uses as well, but they appear to be unaffordable at present. Most of them would also require a significant increase in the size of the NSFG staff. However, they would significantly enhance the NSFG and its usefulness for certain groups of users if they could be done.

Our first priority is to keep the NSFG in the field as close to continuously as is affordable. This will allow us to provide large, representative data sets for our funders, policy-makers, and the research community, while continuing to innovate in both content and survey methodology. For example, if we had funding adequate to continue doing 5,000 interviews per year for 9 consecutive years, we could produce 3 data sets with 15,000 interviews each (45,000 total), and keep questionnaire content responsive to emerging needs over that time period. That would allow studies of trends in the behavior of even small sub-groups of the population that are important for policy and research—such as users of infertility services, fathers who do not live

with their children, the sexual activity of teenagers, marital dissolution at 10-20 years of marriage, childbearing among cohabiting couples, trends in infertility among childless couples, and so on.

We believe that we have made the transition to continuous (or nearly continuous) interviewing successfully. This transition promises to raise the usefulness—and the profile—of the NSFG in the near future.

## REFERENCES

### Reports on the Methodology of the NSFG

1. Groves RM, Benson G, Mosher WD, Rosenbaum J, Granda P, Axinn W, Lepkowski J, Chandra A. 2005. Plan and operation of the 2002 National Survey of Family Growth. Vital and Health Statistics Series 1, No. 42. Hyattsville, MD: National Center for Health Statistics. August, 2005. 86 pages.
2. Groves RM, Mosher WD, Lepkowski JM, Kirgis NG. 2009. Planning and Development of the Continuous National Survey of Family Growth, 2006-2008. Vital and Health Statistics, Series 1, No. 48. NCHS. Sept 2009. 64 pages.
3. Kelly JE, Mosher W, Duffer A, Kinsey S. 1997. Plan and Operation of the 1995 National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 36. National Center for Health Statistics. NCHS. October, 1997. 89 pages.
4. Lepkowski J, Mosher W, Davis K, Groves R, Van Hoewyk J, Willem J. 2006. National Survey of Family Growth, Cycle 6: Sample Design, Weighting, Imputation, and Variance Estimation. Vital and Health Statistics, Series 2, No. 142. NCHS, Hyattsville, MD, July, 2006. 82 pages.
5. Lepkowski J, Mosher W, Davis K, Groves R, Van Hoewyk J. 2009. Continuous National Survey of Family Growth: New Concepts for Sample Design and Analysis. Vital and Health Statistics, Series 2, No. xxx, forthcoming in 2009 or early 2010.
6. Groves RM, Heeringa SG. 2006. Responsive Design for Household Surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society A 169, Part 3: 439-457, April.
7. Martinez GM, Marin BV, Schoua-Glusberg A. 2006. Translating from English to Spanish: The 2002 National Survey of Family Growth. Hispanic Journal of Behavioral Sciences 28 (4): 531-545.

NCHS Reports of Findings from the 2002 (Cycle 6) NSFG

8. Mosher WD, Martinez G, Chandra A, Abma J, Willson S. 2004. Use of Contraception and Use of Family Planning Services in the United States, 1982-2002. Advance Data No. 350, Dec 10, 2004. National Center for Health Statistics. 36 pages.
9. Abma JC, Martinez G, Mosher W, Dawson B. 2004. Teenagers in the United States: Sexual Activity, Contraceptive Use, and Childbearing, 2002. Vital and Health Statistics, Series 23, No. 24. December 2004. 48 pages.
10. Mosher WD, Chandra A, Jones J. 2005. Sexual Behavior and Selected Health Measures: Men and women 15-44 Years of age in the US, 2002. Advance Data No. 362. Hyattsville, MD: NCHS. Sept 15, 2005. 56 pages.
11. Chandra A, Martinez GM, Mosher WD, Abma JC, Jones J. 2005. Fertility, Family Planning, and Reproductive Health of US Women: Data from the 2002 National Survey of Family Growth. Vital and Health Statistics, Series 23, Number 25. December, 2005. Hyattsville, MD: NCHS. 174 pages.
12. Martinez GM, Chandra, A, Abma JC, Jones J, and Mosher WD. 2006. Fertility, Contraception, and Fatherhood: Data on Men and Women from the 2002 National Survey of Family Growth. Vital and Health Statistics, Series 23, Number 26. May, 2006. Hyattsville, MD: NCHS. 142 pages.
13. Anderson JE, Chandra A, Mosher WD. 2005. HIV Testing in the United States, 2002. Advance Data Number 363. Hyattsville, MD: National Center for Health Statistics. November 8, 2005. 32 pages.
14. Anderson JE, Mosher WD, Chandra A. 2006. Measuring HIV Risk in the US Population aged 15-44: Results of the 2002 NSFG. Advance Data No. 377. October 23, 2006. Hyattsville, MD: NCHS. 28 pages.
15. Ventura SJ, Abma JC, Mosher WD, Henshaw SK. 2008. Estimated Pregnancy Rates by Outcome for the United States, 1990-2004. National Vital Statistics Reports 56(15):1-28. Hyattsville, MD: National Center for Health Statistics. April 14, 2008. Available at: [http://www.cdc.gov/nchs/data/nvsr/nvsr56/nvsr56\\_15.pdf](http://www.cdc.gov/nchs/data/nvsr/nvsr56/nvsr56_15.pdf).
16. Jones J. 2008. Adoption Experiences of women and Men and Demand for children to Adopt by Women 18-44 years of age in the United States, 2002. Vital and Health Statistics, Series 23, No. 27. 36 pages.
17. Goodwin P, McGill B, Chandra A. 2009. Who Marries and When? Age at First Marriage in the United States. NCHS Data Brief No. 19. June 2009. 8 pages.

References by non-NCHS authors

18. Finer LB and Henshaw SK. 2006. Disparities in Rates of Unintended Pregnancy in the United States, 1994 and 2001. Perspectives on Sexual and Reproductive Health 38(2): 90-96, June 2006.
19. Kost K, Singh S, Vaughan B, Trussell J, Bankole A. 2008. Estimates of Contraceptive Failure from the 2002 National Survey of Family Growth. Contraception 77 (1): 10-21. January, 2008.