CDC

Monitoring the

Nation's Health

Vital and Health Statistics

August 2005

Series 1, Number 42

Plan and Operation of Cycle 6 of the National Survey of Family Growth

Suggested citation

Vital and Health Statistics

Series 1, Number 42

Plan and Operation of Cycle 6 of the National Survey of Family Growth

Data from the Programs and Collection Procedures

National Center for Health Statistics

Edward J. Sondik, Ph.D., *Director*

Jennifer H. Madans, Ph.D., *Acting Co-Deputy Director*

Michael H. Sadagursky, *Acting Co-Deputy Director*

Jennifer H. Madans, Ph.D., *Associate Director for Science*

Edward L. Hunter, *Associate Director for Planning, Budget, and Legislation*

Michael H. Sadagursky, *Associate Director for Management and Operations*

Lawrence H. Cox, Ph.D., *Associate Director for Research and Methodology*

Margot A. Palmer, *Director for Information Technology*

Margot A. Palmer, *Acting Director for Information Services*

Linda T. Bilheimer, Ph.D., *Associate Director for Analysis, Epidemiology, and Health Promotion*

Charles J. Rothwell, *Director for Vital Statistics*

Jane E. Sisk, Ph.D., *Director for Health Care Statistics*

Jane F. Gentleman, Ph.D., *Director for Health Interview Statistics*

Clifford L. Johnson, *Director for Health and Nutrition Examination Surveys*

Division of Vital Statistics

Charles Rothwell, *Director*

James A. Weed, Ph.D., *Deputy Director*

Stephanie J. Ventura, M.A., *Chief, Reproductive Statistics Branch*

Nicholas F. Pace, *Systems Programming and Statistical Resources Branch*

Acknowledgments

Contents

Appendixes

Text Figures

Text Tables

Abstract

Objectives

This report describes how Cycle 6 of the National Survey of Family Growth (NSFG) was designed, planned, and implemented. The NSFG is a national survey of women and men 15–44 years of age designed to provide national estimates of factors affecting pregnancy and birth rates; men's and women's health; and parenting. Cycle 6, conducted in 2002, was the first time the NSFG included a sample of males.

Methods

The survey used in-person, face-to-face interviews conducted by trained female interviewers. One person per household was interviewed from a national area probability sample in about 120 sample areas, with oversamples of teenagers, African Americans, and Hispanics. The data collection used computer-assisted personal interviewing (CAPI). Separate questionnaires were used for female and male respondents. The last section of the questionnaires used a technique called audio computer-assisted self-interviewing (ACASI). In order to control costs and nonresponse errors, survey managers statistically analyzed results from interviewers' visits to sampled households each day, and used those results to allocate interviewer labor and other resources more efficiently. This management improved response rates and made the sample more representative.

Results

Over 12,500 interviews were completed, about 7,600 with females and about 4,900 with males. The response rate was about 80 percent for females and about 78 percent for males. The survey procedures were adapted during the fieldwork to achieve the desired response rates and to control costs.

Keywords: *Computer-assisted personal interviewing (CAPI) • survey methodology • response rates • audio computer-assisted self-interviewing (ACASI) • eligibility rate • National Survey of Family Growth*

Plan and Operation of Cycle 6 of the National Survey of Family Growth

By Robert M. Groves, Ph.D., and Grant Benson, Ph.D., Institute for Social Research, University of Michigan; William D. Mosher, Ph.D., Division of Vital Statistics, National Center for Health Statistics; Jennifer Rosenbaum, Peter Granda, Ph.D., William Axinn, Ph.D., and James Lepkowski, Ph.D., Institute for Social Research, University of Michigan; and Anjani Chandra, Ph.D., Division of Vital Statistics, National Center for Health Statistics

Executive Summary

This report documents how Cycle 6 of the National Survey of Family Growth (NSFG) was planned and carried out. This kind of information should be useful to at least two types of readers. First, data analysts who intend to use the NSFG, and wish to know more about how and why the survey was conducted the way it was, will find background information that will be helpful in using the data and interpreting the findings. Second, the report is also meant to be useful to those who are interested in survey methodology, and whose surveys might benefit from the approaches used in Cycle 6 of the NSFG.

The report begins with an introduction to the NSFG, including a brief summary of the planning and development that led to the basic design features of Cycle 6 of the NSFG. Cycle 6 of the NSFG is part of a series of national surveys since 1955 that have collected nationally representative data on factors related to birth and pregnancy rates in the United States. Cycle 6 was, however, the first of these surveys to include a national sample of men as well as a sample of women, and it was the largest survey in the series. To incorporate the inclusion of men, a wide range of consultations were made with experts in survey methods and in the topics that the NSFG covers, to determine how the questionnaire for men should be organized, and what topics should be collected. Many of those consulted are listed in the report.

The report then presents the first published summary of the results of the NSFG Cycle 6 Pretest, which tested a number of new ideas for improving data collection, using randomized experiments. As a result of the pretest, a number of important changes were made for the main study: the length of the male questionnaire was reduced to 60 minutes and the female questionnaire to 80 minutes; the interviewer training program was revised to spend more training time on the questionnaires; $40 incentives were used instead of the $20 incentives used in Cycle 5, and interviewers were trained on how to answer respondents' concerns about survey participation.

The next section gives a brief overview of the **sample design** for the NSFG Main Study, which was a multistage national area probability sample, with oversamples of black, Hispanic, and teenage respondents drawn from 121 primary sampling units (PSU's) across the United States. Next, the **questionnaire development process** is summarized, describing the changes necessary to accommodate the inclusion of men into the survey for the first time. The questionnaires were programmed in a software system called BLAISE®, a frequently used system for programming

complex questionnaires. This account includes a description of the efforts made to translate the questionnaires into Spanish and to make such translations culturally appropriate to the major subgroups of Spanish-speaking respondents. The **interviewer materials** used in the NSFG main study are described, followed by an account of the challenges faced in **training and preparing interviewers** for this complex assignment. In brief, the interviewers were trained in in-person sessions for about a week, using small-group techniques for the most difficult topics. Sessions included background on the study, practice answering questions that respondents often have, practice obtaining informed consent, mock interviews going through the male and female questionnaires, and instruction on communicating data back to the contractor's home office.

The next section, **"Data Collection,"** describes the field organization of the study, as well as the new "responsive design" procedures and techniques that were used to monitor and supervise the data collection. These procedures allowed project staff to monitor very closely, on a daily basis, the costs and labor being expended on interviewing, determine what areas and interviewers were doing well or poorly, and predict what sample segments would have the most respondents who would agree to complete the interview promptly. In addition, this section describes the Phase 3 sample, which was carried out during the last month of fieldwork. A sample of the remaining cases was contacted using modified procedures and enhanced incentives. This Phase 3 sample lifted response rates to reach the project's goals, accomplishing a 79 percent overall response rate—including 80 percent for women, 78 percent for men, and 80 percent for teenagers, while controlling costs.

A brief overview of the **quality control** procedures includes efforts to improve quality at training, and a discussion of verification procedures. This is followed by an account of data processing, which resulted in public-use data files containing three subfiles: a female respondent file, a female

pregnancy file, and a male respondent file. The data files were checked selectively, focusing primarily on a few hundred "recoded" variables that NCHS (and many other data users) uses for most reports and articles based on the data. Missing or inconsistent data on these "recoded variables" were imputed—primarily by multiple regression imputation, with some complex cases imputed by logical imputation. The aim of this processing was to produce a data file that NSFG data users could use without further assistance, and to discover problems that should be corrected before the NSFG is done again. The result was a public-use data file that was released on CD-ROM, along with two reports in December 2004. The data and documentation for that public-use file are on the NSFG Web site at: http://www.cdc.gov/nchs/nsfg.htm.

This report also contains a variety of additional information in the Appendixes. Appendix I is a glossary of technical terms. Appendix II is a detailed outline of the male and female questionnaires. Appendix III contains the generic letters and consent forms that were used in the main study data collection. Appendix IV contains the contact scripts used to introduce the survey. Appendix V is an authorization letter to explain the role of the interviewer to a respondent or anyone in the community who wants to know why she is there. Appendixes VI-VIII show the Question-and-Answer Brochure, the Confidentiality Brochure, and the Family Fact Sheet that were used in the field. Appendix IX is the Life History Calendar used in the female interview. Appendix X is an outline of the interviewer training program used for the main study in 2002.

This report focuses primarily on the planning and data collection of the NSFG Cycle 6. A companion report, to be published by NCHS in Series 2 of Vital and Health Statistics (1) will focus on the statistical aspects of the survey—the sample design, sampling weights, imputation, and variance estimation of Cycle 6 of the NSFG.

Planning and Development of Cycle 6 of the National Survey of Family Growth

Introduction to the National Survey of Family Growth

The National Survey of Family Growth (NSFG) is a principal source of U.S. national estimates of factors affecting pregnancy and birth rates, including sexual activity, cohabitation, marriage, divorce, contraceptive use, miscarriage and stillbirth, infertility, and use of medical services for family planning and infertility.

Cycle 6 of the NSFG is one of a series of surveys that began in 1955, based on national probability samples of women 15–44 years of age in the United States. The original purpose of the surveys was to collect data from national samples of women that would help to explain trends and differentials in birth and pregnancy rates. The NSFG still serves this function, but also provides data needed by other Federal programs such as:

- Programs of the Office of Population Affairs concerned with teenage sexual activity and pregnancy, and the use of Title X Family Planning Services
- The Healthy People 2000 and 2010 programs, which use the NSFG for their objectives on Family Planning, HIV Prevention, and Sexually Transmitted Diseases (STDs)
- The National Institute for Child Health and Human Development (NICHD) uses NSFG data to inform and shape its extramural (grant) research programs related to fertility, and to provide a data resource for private and university-based researchers
- The HIV Prevention program of the Centers for Disease Control and Prevention (CDC) uses NSFG data

to provide reliable information on the sexual and drug-related behaviors that increase the risk of HIV and STD transmission.

A Brief History 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. Before then, smaller national surveys of married women were conducted by private organizations in 1955 and 1960. In 1965 and 1970, they were conducted by university researchers with Federal funding (2,3).

As shown in figure 1, the NSFG has been conducted six times since 1973 by NCHS. Picking up where previous studies left off, the NSFG continued making improvements in the national measurement of fertility. 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 interviewing (or most of the interviewing) was done. Cycle 1 was conducted in 1973 and interviewed nearly 10,000 women 15–44 years of age, 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 (4).

NSFG Cycle 3 in 1982 expanded the sampling frame to include all women 15–44 years of age 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 (5). NSFG Cycle 4, fielded in 1988, responded to important fertility and contraceptive questions of the day, including more 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 (6).

In response to recommendations that the NSFG should increase the number and depth of measures used to predict fertility-related variables, several changes were introduced in Cycle 5 (7). A file of contextual data was created, allowing researchers to examine the ways in which characteristics of the place of residence—census tract, local, or State—influence behaviors (8). 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, Computer-Assisted Personal Interviewing (CAPI) on laptop

Cycle	Year	Scope in years	Number of interviews	Source of sample	Over-samples	Average length in minutes	Incentive payment
1	1973	Ever-married women 15-44	9,797	Independent 101 PSUs	Black women	60	No
2	1976	Ever-married women 15-44	8,611	Independent 79 PSUs	Black women	60	No
3	1982	All women 15-44	7,969	Independent 79 PSUs	Black women and teens	60	No
4	1988	All women 15-44	8,450	NHIS 156 PSUs	Black women	70	No
5	1995	All women 15-44	10,847	Independent 198 PSUs	Black and Hispanic women	100	$20
6	2002	Men 15-44 Women 15-44	*12,571	Independent 121 PSUs	Black and Hispanic persons and teens	M=60 F=85	$40

*In 2002 the National Survey of Family Growth surveyed 7,643 women and 4,928 men.
NOTE: PSU is primary sampling unit. NHIS is National Health Interview Survey.

Figure 1. History of the National Survey of Family Growth

computers replaced the paper and pencil questionnaire, improving both the quality and timeliness of the data. Cycle 5 NSFG's 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 was another innovation.

Cycle 6 of the NSFG yielded over 12,500 interviews with men and women 15–44 years of age in the household population of the United States. The project was designed to produce national estimates of characteristics relating to:

1. Trends and differentials in birth and pregnancy rates
2. Determinants of birth and pregnancy rates, including sexual activity, contraceptive use, infertility, and sterilization
3. Marriage, divorce, cohabitation, and adoption
4. Use of medical services for birth control, infertility, and selected health screening and behavior related to the risk of HIV and other sexually transmitted diseases
5. Men's roles in raising and supporting their children
6. Men's and women's attitudes about marriage, children, and families

A Brief Chronology of the National Survey of Family Growth Cycle 6

The initial design work for Cycle 6 of the NSFG attempted to specify the goals of the survey, by answering questions such as the following: If the NSFG includes men,

- What is a realistic response rate for men and for women?
- What topics will yield high-quality data?
- What topics are most needed by policymakers and academic researchers?
- Thus, what topics should be collected?
- How can all this be done without compromising the survey of women?
- How can it be done for an affordable cost?

The NSFG staff awarded a number of contracts to directors of other surveys and researchers in particular subject matter areas. These experts prepared reports and discussed with NSFG staff their recommendations for Cycle 6. Freya Sonenstein, Ph.D., and Laura Porter, Ph.D., then of the Urban Institute in Washington, DC, joined the NSFG staff temporarily to assist in assessing all these recommendations and to begin preparing draft questionnaires for men and for women. These draft questionnaires were then discussed with the representatives of collaborating agencies and other outside experts.

With these topics and sample size recommendations in mind, the NSFG staff drafted the Request for Proposals (RFP) for the contract for Cycle 6, with the assistance of the Office of Research and Methodology, NCHS, and CDC contracts staff in Hyattsville, MD. Next, draft questionnaires and specifications from the draft RFP were included in the protocol that was submitted to the NCHS Institutional Review Board (IRB) in February 1999.

In December 1999, the contract for Cycle 6 of the NSFG was awarded to the University of Michigan Institute for Social Research (ISR). On May 18, 2000, the NCHS IRB approved the Protocol for the NSFG Cycle 6 Pretest. On October 17, 2000, clearance was granted by the Office of Management and Budget (OMB) to conduct the NSFG Cycle 6 Pretest. Pretest data were collected in 2001.

Under contract with NCHS and under the direction of the NCHS staff, the ISR implemented the sample design, constructed the CAPI instrumentation, hired and trained interviewers, administered the data collection, and prepared the data files. It culminated in the creation of data files for public use. An overview schedule of the Cycle 6 activities appears in figure 2, which shows that the first year (2000) was devoted to constructing the questionnaire and sample implementation; the second year (2001), to pretest data collection and analysis and the revision of the survey materials; the third year (2002), to the main data collection; and the fourth and fifth years, (2003 and 2004), to data editing, weighting and imputation, and the

construction of public-use data files and their documentation.

The complicated nature of large-scale household surveys requires a group of researchers, funders, and technical specialists to complete the work. Figure 3 provides a partial list of the principal staff responsible for the design and planning of major aspects of the Cycle 6 survey. Thus, the list in figure 3 does not include all staff that made important contributions by working on particular tasks. Many others were involved for varying periods of time in carrying out the training, interviewing, data processing, and administration of the survey: among those at ISR who played important roles were Karl Dinkelmann, Michael Shove, and John Van Hoewyk.

Important Features of the National Survey of Family Growth Cycle 6

The Inclusion of Males

The most significant innovation in Cycle 6 was the inclusion of a national sample of males 15–44 years of age in the survey. The interviews with males and females contained some topics in common, but also differed significantly in length, complexity, and organization.

New Measurement Based on Expert Reviews of the Cycle 5 Protocol

A number of different studies and reports were sponsored by NCHS after Cycle 5, in an effort to improve the design and content of Cycle 6. These involved a review of the entire questionnaire by survey methodology experts, a review of attitudinal measurements in the NSFG, a review of data on sexual partners, and studies of how to ask questions on nonvoluntary sexual contact, on contraceptive use, on alternative measures of wantedness of pregnancies, and on adoption. In addition, a report on the issues in using ACASI in Cycle 5 was prepared. These studies offered important guidance to improvements in Cycle 6.

1997–99	Commissioned studies on alternative features of the National Survey of Family Growth (NSFG) Cycle 6 by scientists external to the National Center for Health Statistics (NCHS)
December 1999	Start of Cycle 6 contract
January 2000–February 2001	Finalization of content of female and male questionnaires; CAPI, ACASI programming; institutional review board (IRB) approval of human subjects protection procedures; and clearance of survey materials by Office of Management and Budget (OMB)
March–August 2001	Pretest data collection
August 2001–November 2001	Final decisions on main data collection design; IRB approval of human subjects protection procedures; and clearance of survey materials by OMB
February 2002	Staff training sessions for main data collection
March 2002–March 2003	Main data collection
February 2003–October 2004	Data coding, cleaning, and data file preparation

Figure 2. An overview of the chronology of events for Cycle 6 of the National Survey of Family Growth

Contextual Data

Continuing an innovation of Cycle 5, Cycle 6 includes some data that describe the block, neighborhood, and larger social context of the respondent. Some of these data come from the 2000 Decennial Census. Other data come from observations made by listers and interviewers upon first visits to the sample areas. To protect the confidentiality of survey respondents, the contextual data are not part of the public-use file. Qualified researchers who wish to use the contextual data should read reference 8 (8) and then contact NCHS at NSFG@cdc.gov or look at the NSFG web page to learn how to apply for access.

Pretest Survey

Because there were many new ideas for measurement features for Cycle 6, the Cycle 6 Pretest was designed to include a number of randomized experiments to test ideas from the survey methodology literature to improve data collection. Each idea was based on a set of research findings. Randomized experiments tested alternative hypotheses. Each experiment

had a decision rule for determining the choice for Cycle 6. This pretest survey is described in the next section.

Responsive Design

Between Cycle 5 and Cycle 6, response rates for household surveys declined throughout the developed world (9). For this reason, three features of Cycle 6 were introduced to react to that trend. Each of these features is described in more detail later in this report:

- First, during interviewer training, interviewers were given special lessons in dealing with respondents' questions at both the screening stage and the main interview stage (see "Interviewer Training").
- Second, the uncertainties about the amount of effort required to obtain each interview were managed by a set of active quantitative monitoring tools, labeled "responsive design." (see "Data Collection" section).
- Third, monetary incentives (described to respondents as "tokens of appreciation") were increased (see "Data Collection").

The National Survey of Family Growth Pretest

Summary of the Pretest

The NSFG Cycle 6 Pretest was conducted in four geographical sites purposefully chosen to contain three large metropolitan areas in different regions of the country and a small rural county. The pretest employed a two-stage area probability sample design based on 73 segments of city blocks and block equivalents in the first stage, and individual housing units in the second stage. Higher probabilities of selection were assigned to blocks with higher proportions of black (or African-American) residents from the U.S. Decennial Census, with a goal to achieve a pretest sample in which about 25 percent of all respondents were black persons.

A total of 2,403 listed addresses were sampled and screened to locate age-eligible persons. Households without English-speaking adults were eliminated from the sample because the pretest was

National Survey of Family Growth (NSFG) staff, National Center for Health Statistics (NCHS)

(In alphabetical order; current staff marked with an *)

*Joyce C. Abma, Ph.D.   *William D. Mosher, Ph.D.
Matthew D. Bramlett, Ph.D. Laura Porter, Ph.D.
*Anjani Chandra, Ph.D.   Stephanie J. Willson, Ph.D.
*Gladys Martinez, Ph.D.

Chief, Reproductive Statistics Branch: Stephanie J. Ventura, M.A.

Design Statistician: Karen E. Davis, M.A.

Office of Population Affairs: Eugenia Eckard, M.A.

National Institute for Child Health and Human Development (NICHD)

Christine Bachrach, Ph.D., and Lynne Casper, Ph.D.

U.S. Department of Health and Human Services, Children's Bureau, Administration for Children and Families

Penelope Maza, Ph.D.

U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation

Kelleen Kaye, M.P.P., and Linda Mellgren, M.P.A.

Centers for Disease Control and Prevention, Division of HIV/AIDS Prevention

John E. Anderson, Ph.D.

Outside experts

Larry Bumpass, Ph.D., University of Wisconsin, Madison, WI
Jacqueline E. Darroch, Ph.D. Alan Guttmacher Institute, New York, NY
David Fine, Ph.D., Center for Health Training, Seattle, WA
Wendy D. Manning, Ph.D., Bowling Green State University, Bowling Green, OH
Kristin A. Moore, Ph.D., Child Trends, Inc., Washington, DC
Freya Sonenstein, Ph.D. The Urban Institute, Washington, DC
Dawn Upchurch, Ph.D., UCLA School of Public Health

Principal University of Michigan staff involved in NSFG design and planning

William Axinn, Ph.D.   Robert Groves, Ph.D.
Erik Austin, M.A.   Lynette Hoelter, Ph.D.
Anne Biddlecom, Ph.D.   James M. Lepkowski, Ph.D.
Grant Benson, Ph.D.   Patricia Maher
Mick P. Couper, Ph.D.   Roger Tourangeau, Ph.D.
Peter Granda, Ph.D.   Krishna Winfrey

*Staff who contributed to design and planning only. Organizational affiliations are those in effect when the person worked in the National Survey of Family Growth Cycle 6 design and planning and may have changed by the time of publication of this report.

Figure 3. Principal people involved in the design and planning of Cycle 6 of the National Survey of Family Growth

conducted in English. One age-eligible person was selected in each household, with an attempt to obtain a roughly equal number of male and female respondents. The final sample included 941 eligible persons and yielded 615 completed interviews. The pretest included many experimental variations that informed final design decisions of the main survey.

Experimental Features Designed Into the Pretest

The pretest contained a variety of features deliberately introduced to study the coverage, nonresponse, and measurement error properties of the data collection process. Specifically, the following experimental comparisons were made:

1. Short vs. long screener questionnaires to determine age eligibility of household members (to study the potential effect of shorter screeners on coverage, cost, and response rates)
2. Probes for household membership in the roster vs. a roster with no such probes (to obtain more complete counts of young

minority males than is common in surveys)

3. $20 vs. $40 incentive for the main interview (to test the effect of higher extrinsic benefits to participate on response rates and data quality; the survey letters and materials refer to these payments as "tokens of appreciation")

4. Alternative positions in the questionnaire for questions on whether pregnancies were wanted or unwanted (to reduce social desirability biases that may be present in the questions)

5. Text only vs. text and audio CASI (to determine whether reporting of sensitive items would be harmed by eliminating the costly step of building computer voice files for ACASI)

6. Follow-up questions, called random retrospective probes, measuring the perceived meaning of responses to key questions (to measure whether the questions were understood as they were intended)

7. Debriefing questions (to measure the comprehension of some key terms such as "nonvoluntary" and "living together")

8. Instructions before key questions emphasizing their importance (to stimulate recall and accurate reporting)

9. Alternative ordering of responses in attitudinal questions (to measure whether there were order effects)

10. Alternative relative position of attitude and behavior questions (in order to measure possible order effects)

11. An attempt to distinguish whether males "didn't remember" or "never knew" certain characteristics of their former wives or sexual partners (to determine the extent to which males could accurately report some important attributes of their sexual partners)

12. Separate vs. combined questions on male and female contraceptive use at last sex among male respondents (to reduce measurement error on these questions)

In addition to these experimental comparisons, there were some observational study components of the pretest intended to gain insight into how the procedures were working. These included interviewer observations about the informed consent protocol and a post-survey focus group with pretest interviewers.

Pretest Sample Design

The NSFG Pretest sample design was not a probability sample of U.S. households, but rather a probability sample of households in four selected geographical areas, which were chosen to reflect a set of design criteria for the pretest. The four pretest sample areas were drawn from three large urban areas in different geographical regions and a small rural county. These were chosen to provide geographic dispersion of sample locations, racial and ethnic diversity, and a mixture of urban and rural locations.

Within each pretest PSU, segments consisting of block groups were chosen. Initial design specifications proposed a total of 100 segments to be selected. In order to provide adequate testing of operational features of the survey, approximately 10 segments were allocated to the rural county, and the remaining 90 segments equally divided among the three metropolitan areas. Within each PSU, up to four strata of block groups were created: block groups with more than 10 percent of the households with black persons, those with more than 10 percent Hispanic households, those with both more than 10 percent black and 10 percent Hispanic persons, and the remainder.

Sample segments were allocated across these strata to increase the expected number of black households in the sample (a target of 25 percent of the households) and to achieve approximately 20 percent of the households in rural locations. Expected response rates for households and persons were 60 percent in all strata. After examining the available budget and performing power calculations on key pretest experiments, a goal of 1,200 completed interviews was set. Several reserve samples were also selected to protect against lower than anticipated eligibility, unexpectedly low response

rates, or longer than anticipated interviews.

Among the target 100 segments, a subsample of 80 segments was designated for first release, with the remaining 20 segments to be released if response rates and interview length required additional sample households. Finally, household samples within segments were divided into two replicates, an initial release replicate of 80 percent of the chosen addresses and the remaining 20 percent of selected addresses that could be released at a later stage if more sample households were needed.

From the selected sample of segments, 80 of the segments and 80 percent of the selected addresses were released for interviewing at the start of the pretest. This yielded a total of 2,625 addresses. After approximately 2 months of data collection activities, significantly higher than anticipated interviewer attrition substantially reduced the rate of production of interviews. Several of the 80 segments had yet to be visited for screening. A subsample of seven segments was chosen to be withheld from further data collection, leaving 2,403 listed and selected addresses. These addresses contained 941 cooperating households with age eligible persons, and yielded 615 completed interviews. Because this achieved sample size was lower than the original goal of 1,200, some of the pretest experiments were based on fewer cases than expected. This may have led to more nonsignificant findings than would have otherwise been the case.

Interviewer Activities

The Institute for Social Research (ISR) administered a multistep training program to prepare interviewers for the tasks of the NSFG Pretest (February 2001). The quality of the data collection was monitored through verification interviews by field management staff on a subset of selected households and selected eligible respondents. Interviewers were organized into four pretest site teams; each interviewer team was led by an interviewer supervisor or team leader. A fifth team leader served as a quality control monitor and completed mock

interviews with selected interviewers to monitor their performance.

Production (screening of households and completion of interviews) started slowly. The field staff experienced a 45 percent interviewer attrition rate in the first 6 weeks of data collection. This attrition rate was largely attributed to a very competitive labor market at the time (March-July 2001), coupled with the complexity and sensitivity of the interviewers' task. Efforts to increase production and retain interviewer staff included frequent interviewer team meetings; visits by team leaders to the pretest sites to work one on one with interviewers in the field; withdrawal of seven segments from the pretest sample; and the addition of supplemental screening staff in two pretest sites. Pretest production ended in the first week of August and yielded 615 completed interviews and 1,791 finalized screeners (exclusive of vacant units and other ineligible cases).

Pretest questionnaire length

The NSFG Pretest involved two complex Computer-Assisted Personal Interview (CAPI) questionnaires—one for males and one for females. These instruments were written in the software system Blaise®, version 4.4 (http://www.westat.com/blaise). This was the first time the male instrument had ever been used in the field. The female instrument drew heavily on the instrument used in previous cycles. Both instruments included an ACASI section of 20 minutes expected length. In addition, a screener was used to implement the complex within-household respondent selection protocol. Each of the three instruments (screener, male and female questionnaires) included a number of experimental questions.

A variety of timing tests were completed during the pretest programming phase. Data on the length of sections of the questionnaire were analyzed to aid in efforts to decrease the length of the questionnaire for the main study.

The mean length of the male questionnaire in the pretest was 91 minutes, while the mean length of the

Table A. Mean length of interview for completed female and male interviews: National Survey of Family Growth Pretest, 2001, and National Survey of Family Growth main survey, 2003

	Minutes	
Interview	Female	Male
Pretest	126	91
Main survey	85	60

female instrument was 126 minutes (table A). The administration length of the male and female questionnaires increased with the number of children and the number of partners reported. In both questionnaires, the subseries of questions asked for each reported partner or pregnancy was long, and these questions were complicated by the need to match partners and pregnancies.

Given the large discrepancy between the desired and achieved questionnaire lengths in the pretest, large amounts of effort were expended after the pretest to reduce the length of the instrument with minimal damage to the research goals of the NSFG. These efforts achieved their targets—a mean female interview length of 85 minutes and a mean male interview length of 60 minutes in the main survey—while still achieving the most important data-collection objectives of the study.

Pretest Interviewer Training Program

The pretest training program used three types of training, administered across four separate sessions:

1. A session to train supervisors and other training staff
2. A session to train interviewers
3. Two sessions of SurveyTrak training (for use of the sample administration software).

Thus, prior to the data collection period each interviewer received 9 days of training.

Each training session made extensive use of small groups, called communities, with less than 15 interviewer trainees for each lead trainer. For more complex modules, training communities were divided into smaller groups of 6–8 trainees. Within each training community, a lead trainer was teamed with an assistant trainer and at

least one "roving" assistant. A total of 11 staff members served as lead or assistant trainers and 6 staff members served as runners across the four training sessions. Training staff was drawn from the field supervisory team as well as the ISR project management team. With input from NSFG staff, ISR developed a trainee manual and a trainer manual to guide the training sessions.

Evaluation and feedback on the three pretest training sessions were solicited from NCHS observers, from trainees, and from the training staff. In addition, after the pretest data collection, a selected group of interviewers, diverse in interviewing experience and accomplishments, was debriefed on concerns about paper work, training, consent forms, questionnaire difficulties, etc. Based on this feedback, it was decided that the training for the main study should spend less time in lecture classes and more time in practice sessions with the male and female instruments. It was also decided that trainers should be given more time for preparation before interviewer training began, so that more time during the training session could be spent on the instruments.

An additional debriefing was conducted on the refusal aversion experiences of the pretest interviewers. The goal was to identify verbal and nonverbal cues that could be used in training for the main study to better address respondents' concerns. Eleven pretest interviewers with high completion rates were selected to participate in this focus group. More than 50 nonverbal behaviors and cues were identified—many of these were typical to any face-to-face interview request, but some were related to the sensitive character of the NSFG questions. The focus group identified several important training areas for the main study, including refusal aversion

Table B. Interviewer observations on upset respondents: National Survey of Family Growth Pretest, 2001

Observations	Counts of respondents (number)	Percent
Total[1]	465	100.0
Not upset during interview	446	95.9
Upset behaviors during interview	19	4.1
Cry	1	0.2
Became angry	2	0.4
Asked for help	1	0.2
Other	6	1.3
Nothing	9	1.9
Still upset at end of interview	0	. . .

. . . Category not applicable.

[1]National Survey of Family Growth pretest respondents for whom ACASI observation data were collected.

aimed at other household members, protection of the respondent's confidentiality during ACASI administration, and a clearer description of the nature of the ACASI questionnaire to the respondents.

Interviewer Observation Forms

The audio CASI part of the Cycle 6 interview was considerably longer in Cycle 6 than in Cycle 5 (20 minutes compared with 8 minutes in Cycle 5, on average). While respondents were conducting the ACASI component of the interview, interviewers completed an observation form on a Palm Pilot Personal Digital Assistant (PDA). Interviewers answered questions about the behavior of the respondent (such as whether the respondent seemed to understand the questions, whether the respondent was attentive or distracted, whether the respondent ever became upset during the interview, whether there were interruptions during the interview, etc.) as well as circumstances of the interview. After the completion of observations, the data were transferred from the PDA onto the laptop.

This interviewer observation was done in Cycle 6 for three principal reasons: first, observations like these have been collected from interviewers in all previous cycles of the NSFG to provide survey managers with data on the circumstances of data collection, and these were deemed necessary in Cycle 6 as well; second, this form kept the interviewer productively busy while the respondent was filling out ACASI; and third, the form provided data requested by the NCHS IRB about respondent reactions to the interview.

Interview observation forms were completed for 466 of the 615 pretest interviews (table B). Due to technical problems with the PDA's and the software used to capture the data, observations could not be recorded for the other 149 completed interviews (about 25 percent).

At the request of the NCHS IRB, one of the observation questions asked about whether the respondent became upset during the interview, accompanied by two followup questions: how the respondent displayed his or her feelings, and did he or she remain upset by the end of the interview. Table B shows that less than 5 percent of the pretest respondents had to stop the interview for more than a few seconds because the respondent became upset in any way. No minors were reported to exhibit any signs of being upset. There were no obvious patterns for the propensity to be upset across gender or age groups. No pretest respondent remained upset at the end of the interview.

Another set of questions asked how frequently the pretest respondent displayed particular behaviors. The observations presented in table C could be interpreted as measures of respondent comfort or discomfort with the interview. On all the observations, most respondents expressed no concerns with the various aspects of the interview. The observation generating the most concern was the embarrassment related to a question in the interview; 30 percent of the pretest interviews had interviewers noting that the respondent appeared embarrassed at least once during the interview. Confidentiality of the data was the least frequently mentioned concern; only 7 percent of pretest respondents mentioned confidentiality concerns at least once.

In short, given the very sensitive nature of some of the questions, the training of the interviewers to adopt a professional, neutral demeanor and to underscore the important statistical uses of the data appears to have been successful.

Overall, the number of persons displaying concern about the content of

Table C. Percentage of respondents exhibiting various behaviors at least once during the interview: National Survey of Family Growth Pretest, 2001

Concern	All[1]	Gender		Age		Race	
		Male	Female	Minor	Adult	Black	Non-Black
Total	465	236	229	44	421	92	373
				Percent			
Asked why we needed to know the answer to one or more questions	9	11	8	2	10	10	9
Appeared embarrassed by one or more questions	30	32	28	30	30	26	31
Avoided making eye contact	23	27	19	32	22	21	23
Asked whether answers would be kept confidential	7	8	6	9	7	8	7

[1]National Survey of Family Growth pretest respondents for whom ACASI observation data were collected.

Table D. Percent distribution of respondents by type of documentation of informed consent by gender and age: National Survey of Family Growth Pretest, 2001

Consent	Total	Gender		Age in years	
		Female	Male	15–17	18–44[1]
			Percent		
Total	100.0	100.0	100.0	100.0	100.0
Respondent signed before interview.	96.7	97.4	96.1	96.4	96.8
Respondent signed after hearing a few questions	1.8	1.3	2.3	3.6	1.6
Respondent signed after the interview	1.1	1.3	1.0	0.0	1.3
Respondent signed at some other point	0.2	0.0	0.3	0.0	0.2
Interviewer signed.	0.2	0.0	0.3	0.0	0.2
			Number		
Sample size[2]	615	311	304	56	559

0.0 Quantity greater than zero but less than 0.05.

[1]18–44 includes three emancipated minors.

[2]Four consent forms were lost in the mail. Interviewers for these cases supplied additional substitute signed forms. These cases have been excluded from the analyses of timing of the consent form.

the questionnaire or the interviewing procedures was quite small. Results from interviewer observations were later used in the training protocols for the main survey as guidance on how to maintain interviewers' sensitivity to respondents' feelings.

Evaluation of Informed Consent Protocols

The NCHS IRB required that the pretest protocol seek written informed consent from all adult sample persons. For sample persons 15–17 years of age, written informed consent to speak with the minor was obtained from an adult parent or guardian; then written informed assent was obtained from the minor. If the sample person wanted to do the interview but refused to sign the form, the interviewer documented this and conducted the interview. At the end of the interview, the respondent was given another chance to sign. Table D shows that almost all respondents (97 percent) signed the form before the interview began.

Questions on informed consent during verification—In addition to asking interviewers to document how each respondent provided informed consent, a verification subsample was reinterviewed. The verification interview took place several days after the pretest interview and, at the request of the NCHS IRB, included questions about the informed consent process. Observations on respondents' reaction to

the informed consent form suggested the documentation of the informed consent should be as flexible as possible in the main study.

Within-Household Coverage Experiments

The NSFG Pretest included an experiment to test household roster questions. This experiment was designed to learn whether it was possible to obtain more complete listings of household members. Previous research had shown that significant percentages of young minority males are not listed in household rosters. There appear to be two types of this under-reporting: some household members may be deliberately under-reporting for various reasons; others may live in two or more places part time and so may not be thought of as "usually" living in the unit.

For one-half the roster sample, informants were permitted to list household members using their initials, first names, or nicknames rather than their full names. Based on previous research (10), the use of initials was expected to reduce the threat that household members could be identified and thereby encourage a more complete listing. The remaining rosters used a traditional approach in which screener respondents were asked to give full names for all household members. Furthermore, special probes were added to the traditional full names to encourage respondents to give a more complete list

of household members. Table E presents the results from these experiments.

Neither of these attempts to improve within-household coverage—using initials instead of full names, or adding more probes to elicit more household members—yielded statistically significant differences from the traditional approach of asking for the full name of each household member, and not adding further probes. Thus, conventional procedures that administer fewer questions and do not require collecting the full name for the sample household member were used in the NSFG Cycle 6 main study data collection.

Pretest Response Rate Analyses

Screener Sample Disposition

To determine whether household members were eligible for the NSFG interview, a screening interview was attempted with each sample housing unit that contained at least one English-speaking adult. Table F presents the disposition of the 1,979 sampled addresses ("sample lines") with occupied housing units. About 83 percent of the sample households yielded a completed screener interview. As expected, the achieved screener response rate was lower than the rate for a traditional face-to-face screening instrument, mainly because nonrespondent followup was limited.

Table E. Percent of households with different numbers of age-eligible household members and mean number enumerated, by experimental roster condition: National Survey of Family Growth Pretest, 2001

Number of age-eligible persons in household	Household members		Roster questions	
	By names	By initials	Added probes	Traditional questions
	Percent			
Total	100.0	100.0	100.0	100.0
None.	32.8	31.7	32.0	32.6
1	22.9	26.1	24.9	24.1
2	33.1	31.2	31.0	33.3
3	7.2	6.8	8.0	6.0
4 or more	3.9	4.2	4.2	3.9
Mean.	1.3	1.3	1.3	1.3
	Number			
Total households.	691	709	704	696

Table G presents disposition codes for screener households eligible for the main interview. The response rate among eligible persons was 67.6 percent. The major reason for nonresponse was refusal to cooperate: 20.5 percent of eligible households refused to cooperate, accounting for about two-thirds of all nonresponse. Table G also shows that females had higher response rates than males (72.5 percent for females and 63.2 percent for males; $p<.013$). Teenagers had much higher response rates than adults (80.4 percent for teenagers and 65.8 percent for adults; $p<.016$). The types of nonresponse differed by race: black respondents had lower refusal rates than other respondents (13.1 percent compared with 23.1 percent; $p<.001$), but higher noncontact rates (10.6 percent vs. 4.1 percent).

Response Rate Analyses: An Experimental Variation in Screener Length

The final response rate among eligible sample persons is a function of three factors:

- The contact rate for sample households
- The cooperation rate on the screening instrument
- The cooperation rate given an eligible respondent

Design features that could increase contact rates and screener response rates were included in the pretest experiments. Some of the procedures intended to reduce the burden of screening could, however, increase coverage error (i.e., miss eligible persons in sampled housing units), so any design change that would make contact and screening easier or cheaper had to be evaluated to see if it caused more eligible persons to be missed (producing undercoverage in the survey).

The pretest used a randomized experimental design, where roughly one-half of the sample segments were assigned to a short screener, and one-half to a traditional screener. A short screener asked the household informant whether everybody in the household was over 55 years of age. If the answer was negative, the interviewer continued with a collection of a household roster. The traditional screener required a full household listing to determine whether there are any age-eligible persons.

Unfortunately, no large response rate gains were found when the short screener was used. In addition, 7 percent of the households reporting no eligibles in the short screener reported one or more eligibles in the validation re-interviews when a full roster was collected.

Several indicators were used to measure whether the amount of interviewer effort was reduced by the short screener: total number of calls made on sample cases, total number of

hours spent in various activities for a sample case, and the percentage of sample cases that were completed on the first contact with a household member eligible to provide the requested information. The results in table H show that the percent of households completing a screening interview on the first contact is higher for the short screener (81 percent) than for the traditional screener (74 percent), by about 7 percentage points (with an approximate standard error of 2.1 percentage points, not accounting for clustering in the design).

The mean number of visits necessary to complete a screener after the first contact is higher for the traditional screener (0.85) than for the short screener (0.47). However, a better net measure of cost impact is the total interviewer screening hours per successful screener, and this shows slightly more hours for the short screener than for the traditional screener. The traditional screener was recommended for the main study for two main reasons: first, because there was insufficient evidence that the short screener would save much interviewer effort, and second, it resulted in more misclassification error than the traditional screener (about 7 percent of the eligible households).

Response Rate Analyses: An Experiment Concerning the Size of Incentives

The response rate goals in Cycle 6 of the NSFG were 80 percent for females and 75 percent for males. In some past household surveys, higher

Table F. Percent distribution of occupied households by outcome of screener: National Survey of Family Growth Pretest, 2001

Disposition	Percent
Total	100.0
Screener completed.	82.6
Refusal.	8.7
Noncontact	8.1
Other noninterview	0.7
	Number
Sample size.	1,979

Table G. Number of households with one or more persons eligible for the main interview, and percent distribution by disposition, according to gender, age, and race of sample person: National Survey of Family Growth Pretest, 2001

Disposition	Total	Gender		Age in years		Race	
		Male	Female	15–19	20 and over	Black	Non-Hispanic black
				Percent			
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Interview.	67.6	63.2	72.5	80.4	65.8	67.1	67.8
Final refusal	20.5	22.3	18.5	13.4	21.5	13.1	23.1
Noncontact	5.7	6.6	4.9	1.8	6.3	10.6	4.1
Other.	6.8	7.9	4.2	4.5	6.3	9.3	5.0
				Number			
Sample size	916	484	432	112	804	237	679
Chi-square	. . .	10.75	. . .	10.34	. . .	26.43	. . .
P-value	. . .	0.013	. . .	0.016	. . .	0.00	. . .

. . . Category not applicable.

0.0 Quantity greater than zero but less than 0.05.

Table H. Percent of screener interviews completed on the first contact with the sample household by screener type (standard error): National Survey of Family Growth Pretest, 2001

Screener group	Percent successful interviews completed in the first contact	Total interviewer screening hours per obtained screener	Mean number of visits to first contact for successful screeners	Mean number of visits after first contact for successful screeners
	Percent (with standard error)			
Short screener.	80.5 (1.4)	4.8 (1.0)	2.6 (0.082)	0.47 (0.044)
Traditional screener.	73.8 (1.6)	3.3 (0.43)	2.3 (0.076)	0.85 (0.084)

nonresponse rates have been found for males than for females, and among younger persons than among older persons (11,12). An additional challenge to the NSFG Cycle 6 is the disproportionate selection of black and Hispanic respondents, who often have lower cooperation and response rates. Among the various tools to increase cooperation rates, incentives have proven to be particularly successful (13).

An experiment concerning the size of incentives was conducted in the NSFG Pretest. Two incentive sizes were randomly assigned to sample households: $20 (the amount used in Cycle 5) and $40 (the experimental amount). The $40 amount was chosen to be sufficiently larger in size in order to test whether it would help to achieve the response rate goals. The $20 and $40 amounts were described to respondents as a "token of appreciation." The amounts were mentioned in the letters sent to respondents. The letters were identical except for the amount of the "token of appreciation."

Two critical sets of indicators were of interest—nonresponse and cost indicators. Table J shows a significantly higher response rate for the $40 incentive group (72 percent) than for the $20 group (62 percent; $\chi^2 = 13.5602$, $p<.0036$).

Table K shows that almost all demographic groups showed positive reactions to the higher incentive amount. Most dramatic were the effects for black sample persons—a 59 percent interview rate for the $20 incentive compared with 74 percent for the $40 incentive (with 6.1 percentage points standard error of the difference, not accounting for the complex sample design). Similarly, females attained a 62 percent interview rate with the $20 incentive and 81 percent, with the $40 incentive (a 4.3 percentage point standard error on the difference, not accounting for the complex design).

Based on these results, it was concluded that higher cooperation rates are possible with higher incentives. This finding is particularly important given

the fact that the main survey oversampled black persons. The pretest results however suggested no large incentive effects for males and for teenagers—the two groups with traditionally lower response rates— with these incentive amounts.

Prior research shows that higher incentives lead to lower required interviewer effort (14). Table L shows that 11 percent of households offered $20 showed reluctance on the first contact, compared with 4 percent of households offered $40. Partly because of this, the number of interviewer hours per completed interview in the $40 incentive condition seemed reduced, although not statistically different (3.6 hours for the $20 condition vs. 2.9 hours for the $40 condition, not counting interviewer hours for screening activities). Given these differences, it seemed likely that the higher incentive would produce gains in interviewer efficiency that were equal to or greater than its increased cost to the project.

Table J. Percent distribution of sample persons by disposition category, according to incentive amount: National Survey of Family Growth Pretest, 2001

Disposition	Incentive amount (dollars)	
	20	40
	Percent distribution	
Total	100.0	100.0
Interview.	62.4	72.1
Refusal	22.0	19.3
Noncontact	7.7	4.1
Other noninterview	7.9	4.5
	Number	
Sample size	428	488

Questionnaire Performance Analysis

Estimation Procedures

Because of a very tight time schedule for pretest analysis and decision-making for the main study, a strategy for variance estimation was developed that allowed analysis of the pretest results to proceed quickly across a large number of staff without requiring variance estimation that accounted for strata (PSUs) and clusters (segments) in every run. Analyses in which entire segments were allocated to an experimental group (for example, the screening and incentive experiments) used variance estimation that accounted for stratification by PSU and clustering by segments. Design effects for these analyses varied, depending on the outcome measure of interest. For example, in the analysis of the screener experiment, household cooperation rates were compared between short and long screener groups. The design effects for the cooperation rates were 2.24 and 2.25, respectively. Given these relatively small values of the design effect in analyses involving perfect homogeneity of the treatment group within segments, it was decided that more expeditious analyses could be completed without accounting for the sample design in every run. All analyses on the pretest data, including the computation of test statistics, are based on simple random sampling assumptions; thus, if the reader is interested in inference to the combined populations of the sample areas, the standard errors are underestimated.

Sensitive Questions

Response distributions on questions about abortion, number of sex partners, and age at first sexual intercourse were examined. The percentage of women who reported abortions showed overall little difference by mode (interviewer-

Table K. Percent distribution of sample persons by disposition, according to selected characteristics: National Survey of Family Growth Pretest, 2001

Characteristic	Disposition					Total eligible
	Total	Interview	Refusal	Noncontact	Other noninterview	
	Percent distribution					Number
Gender						
Male (incentive):						
$20	100.0	62.3	20.4	7.4	10.0	231
$40	100.0	63.6	24.4	6.0	6.0	250
Female (incentive):						
$20	100.0	62.1	24.1	8.2	5.6	195
$40	100.0	80.6	14.2	2.2	3.0	232
Age						
15–19 years (incentive):						
$20	100.0	78.7	12.8	2.1	6.4	47
$40	100.0	81.3	14.1	1.6	3.1	64
20–44 years (incentive):						
$20	100.0	60.2	23.2	8.4	8.2	379
$40	100.0	70.3	20.3	4.6	4.8	418
Race						
Black (incentive):						
$20	100.0	59.1	12.7	15.5	12.7	110
$40	100.0	73.8	13.5	6.4	6.4	126
White and all other (incentive):						
$20	100.0	63.3	25.3	5.1	6.3	316
$40	100.0	71.1	21.6	3.4	3.9	356
Total (incentive):						
$20	100.0	62.2	22.1	7.8	8.0	426
$40	100.0	71.8	19.5	4.2	4.6	482

Table L. Percent of main interview eligible respondents exhibiting reluctance and total interviewer hours per completed interview by incentive condition (standard error): National Survey of Family Growth Pretest, 2001

Incentive group (dollars)	First contact reluctance	Total interviewer hours per obtained interview
	Percent (standard error)	Mean (standard error)
20	11.1 (1.6)	3.65 (0.54)
40	3.9 (0.9)	2.90 (0.27)

administered vs. ACASI). The $40 incentive, however, appeared to be associated with higher (more complete) reporting of abortions among black women, as shown in table M.

Thus, table M shows that 45 percent of black women and 22 percent of white and other women, reported having had an abortion when they were given a $40 incentive. The figures were 32 percent of black and 23 percent of white and other women, using a $20 incentive. Abortion rates in the United States in 2000 were 57 per 1,000 (5.7 percent) for black women and 21 per 1,000 for all women—a ratio of more than 2 to 1 (15) (table 1 of reference 15). The proportion ever having had an abortion is not the same statistic as the abortion rate, but this comparison does suggest that higher reporting is better reporting, and that abortion reporting for black women was better in the pretest for women who got $40 than for those who received $20.

Consistent with other data, men reported more sex partners than women in both modes (interviewer-administered vs. ACASI). There were upward shifts in the distributions of number of sex partners for both men and women in ACASI. There was no consistent pattern between higher incentives and higher reports of sex partners in either mode or for either gender.

Question Placement for Wantedness of Pregnancies

NSFG Cycle 5 placed contraception questions before questions on pregnancy wantedness (whether the respondent wanted another child and wanted a child at a given time). This created a potential for reporting bias, since admitting that one did not use or stopped using

contraception might have led to positive responses on the wantedness of pregnancy questions. Also there was a potential problem of consistency bias; that is, respondents trying to match attitudes with prior reported behaviors. This could lead to fewer reports of ''unintended'' pregnancies among those using contraception and fewer reports of ''intended'' pregnancies among those using contraception prior to pregnancy.

Because of these concerns, a pretest experiment tested whether placement of attitude questions about wantedness of a pregnancy before or after behavior questions about contraceptive use makes a difference in reports on wantedness of pregnancies. No statistically significant associations between question placement and reports of pregnancy wantedness were found, so the question order was kept consistent with previous cycles.

Audio CASI versus Text CASI

While the survey field has rapidly embraced audio-CASI for the self-administration of sensitive items (16,17), there is as yet little evidence regarding its advantages over text-CASI (computer

assisted self-interviewing with no audio). Given the added effort, cost, and computer capacity required to conduct an audio-CASI interview, it was considered important to evaluate its usefulness in a large sample. In addition, audio-CASI may decrease the standardization of the measurement by giving the respondent a choice of whether to listen to the audio questions, or to read them on the screen.

One belief related to the advantages of audio-CASI is that it reduces nonresponse and measurement errors among respondents with low literacy. A second is that the use of headphones increases respondent's perception of privacy, eventually inducing more honest answers.

A related issue with audio-CASI (and with any self-administered instrument) is whether the respondent needs assistance to complete the self-administered questionnaire. This problem has never been systematically measured for audio-CASI instruments. It was important to find whether audio-CASI reduces nonresponse relative to text-CASI.

In the NSFG Pretest the male and female samples were split into half samples and each half was randomly assigned to audio-CASI or text-CASI administration. No consistent differences in response distributions between text-only and audio and text computer self-administration were found. The average time necessary to complete the self-administered portion of the interview was longer only for the male version of the audio-CASI instrument.

Table M. Percent of women 15–44 years of age who reported that they have had one or more abortions in their lifetime, by race and amount received: National Survey of Family Growth Cycle 6 Pretest

Race	Total	Amount received (dollars)	
		20	40
		Percent	
All women	28	26	30
Black women.	40	32	45
White and all others	22	23	22
t-test: Black/white and all others	0.01	NS	0.01
Ratio: Black/white and all others.	NA	1.4	2

NA Not available.
NS Not significant.

Overall, the results suggested that audio-CASI might not yield consistently superior quality data relative to text-CASI. However, given the importance of the NSFG estimates and the usefulness of audio-CASI for illiterate respondents, the audio version of the computer self-interview was recommended for the main study.

Random Retrospective Probes

Past research has demonstrated the value of asking a random subsample of respondents to elaborate their answers to selected closed questions. To implement the random probe technique in the pretest, a random number identified a subset of interviews, which received an additional probe for a short set of predesignated questions. Such "random probes" can identify comprehension or recall problems and facilitate wording changes (18).

Because male fertility and family dynamics are of central importance in the NSFG Cycle 6, this question was targeted for a random probe with the goal of increasing men's reporting of pregnancies they have fathered. For example, a random probe may elicit events that were not stored in memory as pregnancies, because they did not end in a live birth. To the extent that such probes are successful in generating additional reports of pregnancies, they can suggest the need for revisions to the final instrument.

Of the 306 male interviews in the pretest, 107 were exposed to the random probes, with a total of 111 probes proffered by interviewers. None of the probes led to a report of a pregnancy not mentioned earlier. As a result, structured probes were not used in the main study.

Debriefing Questions as Indicators of Comprehension Problems

Debriefing questions are queries posed to the respondents after they provide an answer to a focal question. Debriefing questions can be used to generate empirical estimates of the proportion of respondents who do not understand the focal question task as intended.

Four topics or questions were chosen to be investigated with these "debriefing" questions: Debriefing questions at the end of each instrument queried the respondent about the meaning of (a) "nonvoluntary" intercourse, (b) "living together," (c) the "current" sexual partner, and (d) features of the Life History Calendar to identify the "last 12 months" before the interview.

The first debriefing question probed the use of the term "living together." Most respondents described "living together" as living in the same house (respondents used phrases such as "in one place," "sharing a home," or "under the same roof"). The majority of respondents described a "current" sexual partner literally—words like "now," "today," or "present time" referring to a sexual partner were often used.

The second debriefing question asked respondents to say whether intercourse was "voluntary" or "nonvoluntary" when (a) alcohol or drugs were given; (b) one person was smaller or younger than the other; (c) the person was told that their relationship would end unless they had intercourse; (d) one was pressured by the other person's words or actions; or (e) one person threatened to harm the other. The majority of respondents thought sexual intercourse was "nonvoluntary" when threats of harm were used. The proportions of men and women who viewed intercourse as "nonvoluntary" were not significantly different, suggesting that men and women in the pretest interpreted these terms in similar ways.

The final debriefing question asked respondents to look at the Life History Calendar and identify the "last 12 months." The vast majority of respondents (92 percent of male and 95 percent of female respondents) correctly identified the last 12 months on the paper calendar.

Based on these results, the meaning of the phrase "living together" was reinforced in the main study questionnaire, and it was recommended that a definition for "current sexual partner" be added in the main study instrument.

Instructions Regarding Question Importance

A typical finding from methodological studies is that socially controversial behaviors, such as abortion or drug use, are underreported. Conversely, socially approved behaviors, such as voting or reading books, tend to be overreported. The social psychological construct most often used to explain these findings is the "social desirability effect;" that is, the tendency for a person to avoid reporting behaviors or attributes believed to be negatively valued by others and to maximize reporting behaviors and attributes believed to be positively valued (19).

Several questions on the NSFG questionnaire could be subject to this type of bias. Examples of behaviors likely to be underreported include abortion, homosexual behavior, and nearly every behavior ascertained in the ACASI sections of the male and female questionnaires. Examples of behaviors that may be overreported include child support payments to nonresident children and contraceptive use for unmarried sexual intercourse.

For some questions in the NSFG Pretest, a prefatory statement was included, emphasizing the importance of the question to the goal of the survey, and asking the respondent to give careful thought to his/her answer. The hypothesis is that increased awareness of the importance of the question, as well as additional time spent formulating the answer, may lead to increased reporting of typically underreported behaviors (e.g., abortion, homosexual behavior, etc.) and decreased reporting of typically overreported behaviors (e.g., child support).

For purposes of the pretest, questions for this experiment were selected based on the importance of the question to the NSFG and its data users and sponsors, other pretest experiments proposed for these items, and distance from other similar introductions in the interview. For women, these topics were contraceptive use, sex with another woman, and income. For men, these topics were fatherhood of children with

former wives/cohabiting partners, child support payments, sex with another man, and income. Introductory statements about the importance of each topic were administered to a random half-sample of men and women in the survey.

With respect to income questions, the experimental introductions were associated with more refusals and "Don't know" responses to some items. However, the majority of income questions did not show any remarkable differences between the experimental and the control group. The experimental introductions were associated with higher percents of women reporting that they received financial assistance for child care (11.5 vs. 4.5 percent)—one of a series of questions on sources of income. Surprisingly, women who received the experimental introductions were less likely to report contraceptive use than women who received the standard introduction. Because there were few large or systematic differences between the groups with and without the introductions, the decision was made to omit the introductions for the main study.

Order of Attitudinal Questions within Section

Male and female respondents in the pretest were asked parallel sets of attitudinal questions, such as:
Please tell me how much you agree or disagree with each statement.
"It is okay for an unmarried 14 year old female to have sexual intercourse.
It is okay for an unmarried 16 year old female to have sexual intercourse.
It is okay for an unmarried 18 year old female to have sexual intercourse."

For each question, respondents were asked:
"Do you strongly agree, agree, disagree, or strongly disagree?"
A similar set of questions was asked, substituting the word "male" for "female."

In the pretest, the order of these was varied in two ways:

1. In some, the questions about an unmarried male were asked first; in others, questions about an unmarried female were asked first, male second.
2. In some, the questions about a 14-year-old were asked first, while in others, questions about an unmarried 16-year-old were asked first; and in still others, an unmarried 18-year-old was asked about first.

Whenever a series of questions asks for similar judgments about members of the same class (such as different subgroups of a larger group like adolescents), question order effects are likely to occur. Such effects present two problems for any survey, including the NSFG—they may (a) distort comparisons between attitudes among males and females, and they could (b) distort efforts to estimate changes in attitudes over time.

For half the sample, the question about males was asked first; for the other half a corresponding question about females was asked first. The items of interest focused on acceptable ages for premarital intercourse for teenage males and females, and three items each about sex roles of males and females.

In addition, the questionnaires systematically varied the order of age presentation. Approximately one-third of the sample began with a question about 14 year olds and moved up in age; another one-third began with the question about 18 year olds and moved down in age, and the remaining one-third was asked about 16 year olds first. The experiment was based on the hypothesis that it might seem more acceptable for a 16-year-old to have sexual intercourse if that item came after a similar question asking about 14 year olds.

No consistent question order effects were detected—there was only one significant result out of 18 comparisons. Based on these findings, it was recommended to ask the questions about females first in order to reduce the risk of affecting responses by a comparison to the male items. Another alternative was to rotate the order of the male and female items in the main study, but for simplicity, the first option was favored.

Although there seemed to be a pattern with higher percentages of respondents agreeing that sex was permissible when the question about 18 year olds was asked first relative to asking about 14 year olds first, the differences were significant only for one item. It was recommended to start with the question about 14 year olds first.

Relative Order of Attitudes and Behavior Sections

A common problem in surveys that mix attitude and behavior measurement is possible order effects of responses—earlier questions affecting responses to later questions. For example, recall of emotional states early in an interview can act as cues to recall of events asked later in the interview; so too, recalling events early in an interview can evoke emotional states that shape responses to later attitudinal questions.

In Cycle 6, an entire section assessing attitudes toward reproductive behaviors was added. This section was placed after a section asking respondents about a wide variety of behaviors. Because of that question order, there was some concern that the attitudinal reports might be biased in the direction of consistency with earlier reported behaviors. However, placing the attitudinal questions first posed other issues. Such an order might threaten comparisons over time with other surveys, where the attitudinal measures are assessed after behavior. More seriously, asking a number of attitude questions about a behavior (such as condom use) could encourage respondents to report that behavior.

To explore such context effects, half of the pretest sample received the attitude section at the beginning of the interview, immediately after the background section, while the other half was administered this section at the end of the interview, just before the ACASI section of the instrument. This was done separately for both male and female respondents.

Overall, there were no consistent directional associations between question placement and attitudes (20 items). Of the 11 behavior items examined, there were also only a few

large differences by placement of the attitude questions—early placement of attitude questions was associated with higher reports of being unmarried when the first child was born (by about 14 percentage points among males) and being at younger age when the first child was born (by about 1.5 years among females). Question placement had different associations by gender for reports of being childless and reports of never being married. Effects on other behavior items were relatively small. Based on these results, for the main study the attitudes section remained as originally placed, near the end of the survey instrument.

Effects of Probes for "Don't Know" and "No" Responses on Male Reporting of Reproductive Events and Behaviors

Past research shows that collecting data from men about their fatherhood and contraceptive use presents some difficulty, particularly when men are not married to their sexual partners. Men tend to under-report nonmarital births, births from previous marriages, and partners' pregnancies, that they have fathered (20). This under-reporting could be because (a) the man can no longer recall something that he knew at one time, or (b) he never knew this fact at all.

An experiment with two goals was conducted in the pretest: to disentangle the reasons for "Don't know" responses by distinguishing between men who once knew something but no longer recall it, versus those who never knew, and to investigate the extent to which men think that events could have occurred without their knowledge. Male respondents were randomly assigned to a control group (no followup probes) or an experimental group, which received followup probes for "Don't know" or "No" responses to questions about contraceptive use, reports of pregnancies and their outcomes (abortion, miscarriage, etc.).

The experimental probes elicited a substantial percentage of responses from male respondents that an event or

behavior could have occurred without their knowledge. Such probes were recommended for the main study.

Question Format on Contraceptive Method at Last Sex

Variation in estimates of contraceptive method used at the most recent sexual intercourse (hereafter, "last sex") may be due to differences in question wording and a variety of other differences across surveys (21). This experiment in the NSFG Pretest compared the estimates of method use at last sex from the male questionnaire using two different question formats:

- A two-question format simultaneously asking about both the respondent and his partner's contraceptive use (the format used in the NSFG Cycle 5)
- A four-question format first asking about respondent's contraceptive use at last sex, then about his partner's method

The two question format was:

"Please look at Card 45. That last time that you had sexual intercourse with ... did you or she use any methods to prevent pregnancy or sexually transmitted disease?"

(If yes): "Still looking at Card 45, that last time, what methods did you and ... use to prevent pregnancy or sexually transmitted disease?"

The four question format was:

"Please look at Card 46. That last time that you had sexual intercourse with ... did you, yourself, use any methods to prevent pregnancy or sexually transmitted disease?"

(If yes): "Still looking at Card 46, that last time, what methods did you, yourself, use to prevent pregnancy or sexually transmitted disease?"

"Please look at Card 47. That last time that you had sexual intercourse with ..., did she use any methods to prevent pregnancy or sexually transmitted disease?"

(If yes) "Still looking at Card 47, that last time, what methods did she use to prevent pregnancy or sexually transmitted disease?"

The four-question format asking separately about the respondent and his partner yielded significantly higher reports of "any method used at last sex" (both for married/cohabiting partners and others). In part, because of a programming error that reduced the sample size for this experiment, however, no statistically significant associations were found between question format types and reporting of the most common individual methods, including condom use and pill use. Because of these inconclusive results, and the importance of this issue for the NSFG, it was decided to repeat this experiment in the main study of Cycle 6.

Effort Analysis for the Pretest

As part of the pretest's preparation for the main survey, the ISR tested systems to capture the costs of key components of the data collection activity. The design of the pretest and main study were significantly different: response rate targets for the main study were higher, resulting in more effort being directed toward more difficult cases; the questionnaires were shorter in the main study; and the fieldwork for the main study was in 120 areas in most States compared with 4 areas in the pretest. Because of these differences, we could not expect the cost structure of the main study to be the same as the cost structure of the pretest. In addition, pretest implementation difficulties such as high interviewer attrition and long interviews limit the utility of the cost and effort data. Despite these differences, the interviewer effort data from the pretest indicated that the original cost estimates for Cycle 6 were likely to be below the actual data collection costs.

Summary of Changes Made to the Main Study as a Result of the Pretest

Based on the pretest results, the following actions were taken for the main study:

1. Questionnaire length was reduced by 30 minutes for males, to 60 minutes, and by 40 minutes for females, to the 80 minute target.
2. The interviewer training program was revised significantly to reflect the complexity of the combined female and male questionnaires.
3. Permission was sought and obtained from the NCHS and University of Michigan IRBs for more flexibility in the informed consent documentation.
4. A traditional household roster screening instrument (with full names) was chosen.
5. Permission was sought and obtained to use a $40 incentive.
6. Refusal aversion training for interviewers, which customized reactions of interviewers to the concerns of respondents, was chosen for the main study.
7. The Cycle 5 placement of contraception and wantedness questions was used.
8. Audio CASI was chosen rather than text CASI.
9. The meaning of "living together" and "current partner" were stated explicitly in the questionnaire.
10. The use of prefatory instructions was not expanded further.
11. The attitudinal questions were placed toward the end of the questionnaire, after most of the key behavioral measures.
12. Probes for males regarding the distinction between "never knew" and "don't recall" were included on a few key questions.
13. An experiment that randomly assigned males to a two-question sequence or a four-question sequence asking about contraceptive use at last sex was included in the main study.
14. On questions about acceptable ages for first sex, females were referred to first, then males, starting with 14 years of age and moving to higher ages.
15. For the laptop computers used by the interviewers, laptop memory was increased from 128K to 256K RAM, and a second battery pack and a longer extension cord were provided to interviewers.

Sample Design

Summary of Sample Design

The sample design of Cycle 6 of the NSFG is described in more detail in reference 1. This section will give a brief summary of the design.

The target population for NSFG Cycle 6 included men and women 15–44 years of age in the household population of all 50 States and the District of Columbia. For Cycles 1–5, the NSFG interviewed only women; thus, men 15–44 were included in the NSFG for the first time in Cycle 6. The NSFG has relied on national probability samples of households and eligible persons for all cycles. In all six cycles, data have been collected from women on trends and changes in fertility, contraception, breast feeding, use of family planning services, and maternal and child health. In Cycle 6, data were also collected from men on fertility, family formation, parenting, and closely related attitudes and opinions.

Cycle 6 employed a stratified multistage probability sample of households and eligible persons. The decision to use this kind of a design was based on the following considerations:

- The target population for Cycle 6 is the household population of women and men ages 15–44 years in the 50 States and the District of Columbia.
- The sample was to consist of approximately 7,500 female and 4,800 male respondents from a probability sample of households. The sample was designed to yield at least 2,000 completed interviews with teens 15–19 years of age, 2,000 with black persons, and 2,000 with Hispanic persons.

- The target interview completion rate for females in the population was to be 80 percent of eligible women and 75 percent of the eligible men. The interview completion rate is the product of household screener response rates and the main interview response rate.
- The stratified area probability design included 110 selected primary areas supplemented by 11 areas chosen to improve the precision of estimates on the Hispanic population.

Area Probability Sample

The NSFG Cycle 6 sample design consisted of four stages of selection to choose eligible sample persons. Women, teen 15–19 years of age, young adults ages 20–24 years, and black and Hispanic persons were selected at higher rates, yielding an oversample of such persons in Cycle 6. The sample design is summarized in figure 4.

Stage 1 of sample selection—The entire land area of the United States is divided into approximately 3,300 counties and independent cities that can be identified through publications of the U.S. Census Bureau. The counties and other units were combined to form a set of 2,402 PSUs for the first-stage selection of an area probability sample of households in the United States.

Two separate samples of U.S. counties were selected for Cycle 6. A national sample selected 110 PSUs from the 50 States and the District of Columbia. The national sample PSU selection was based on 1990 Census counts of occupied housing units in each county or other unit.

This national sample was efficient for obtaining a sample of black, white, and other race households. For the national sample of Hispanic households, another sample was used. Since the selection of the second Hispanic PSU sample was completed before 2000 Census household data were available, the counties were selected using counts of Hispanic persons rather than occupied housing units. The same 2,402 PSUs were divided into 45 strata, from which one PSU was selected. Selection was made with probability proportional to U.S. Census 2000 Hispanic population,
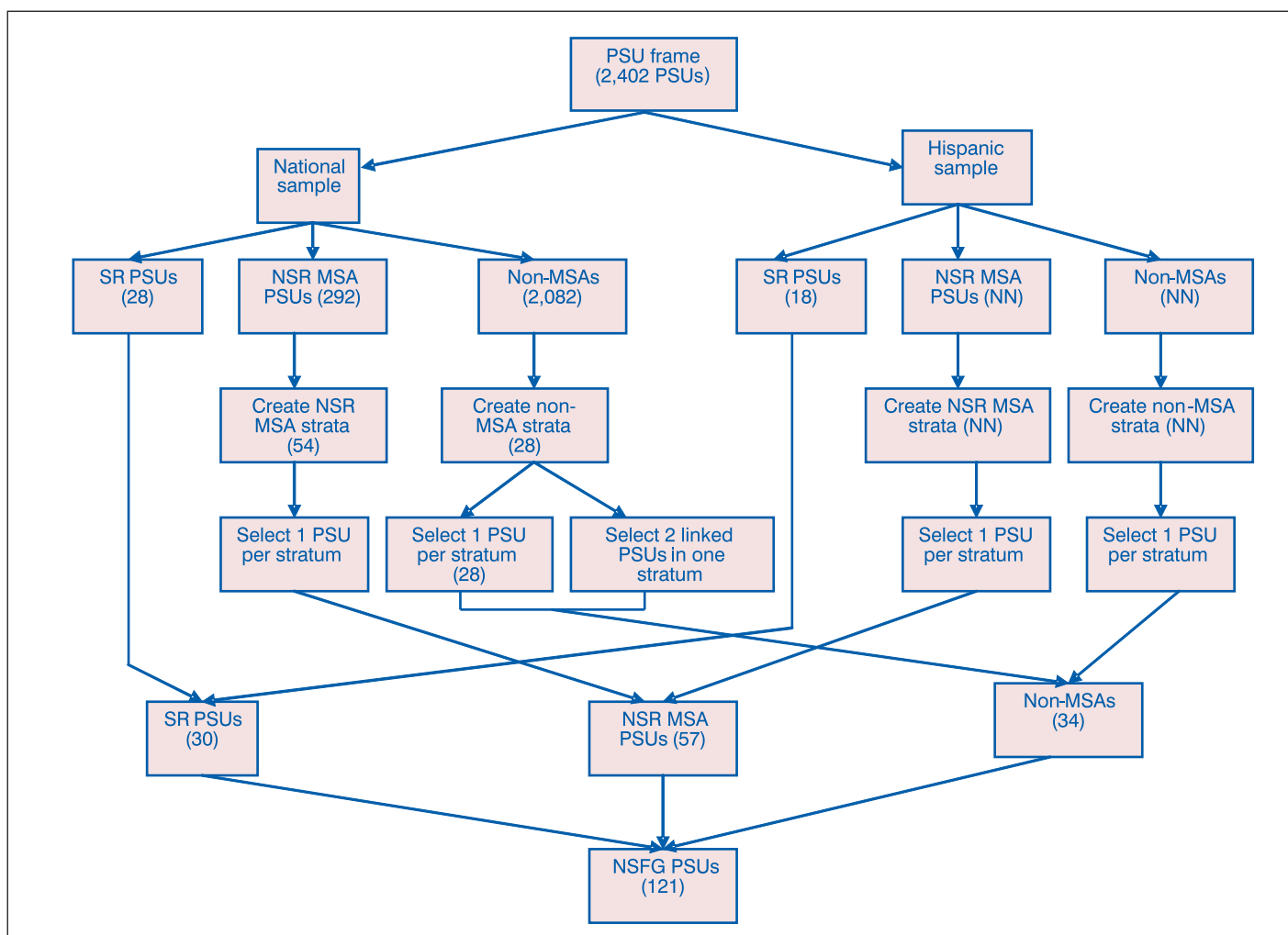
Figure 4. National Survey of Family Growth (NSFG) Cycle 6 selection of primary sampling units and segments

and with a technique designed to maximize overlap in selected PSUs with those in the national sample. The selection yielded a total of 34 Hispanic sample PSUs that were already in the national sample and 11 additional Hispanic PSUs.

Stage 2 of sample selection—Within the 121 national and Hispanic sample PSUs, a list of blocks was obtained from the 2000 Census. Information was available on the population by race and ethnicity. An estimated number of occupied black, Hispanic, and white or other race households were obtained by dividing the year 2000 counts of black, Hispanic, and white or other races persons, by the U.S. average number of persons in households with a black, Hispanic, or other race head.

In the second stage of selection, blocks were divided into four domains within each PSU: (1) nonminority,

(2) more than 10 percent black households, (3) more than 10 percent Hispanic households, and (4) more than 10 percent black and more than 10 percent Hispanic households. Blocks were selected within each domain with probabilities proportionate to the estimated number of households in the block in the 2000 Census.

In order to improve data collection efficiency by reducing travel costs, blocks falling below a minimum size were combined with other geographically contiguous blocks. In urban areas, the minimum size was 75 estimated households, while in rural areas that minimum was 50. The final units consisted of entire blocks or sets of linked blocks, and are referred to as segments.

Stage 3 of sample selection—Trained household listers were sent to each of the 1,414 sample segments

selected in the second stage to list housing units on the blocks in the segments. Housing unit lists were returned to the Survey Research Center and keyed into a file containing one line per housing unit.

The third stage of selection chose sample lines (housing units) from each sample segment with probabilities designed to obtain equal chances of selection for households within segment domains. Lines in domains 2, 3, and 4 (figure 5) were selected at higher rates to increase the number of black and Hispanic households in the sample.

Stage 4 of sample selection—The fourth stage was the selection of eligible persons within sample households. Interviewers visited housing units represented by each selected sample line. When the housing unit was occupied, interviewers attempted to list persons. One eligible person was chosen
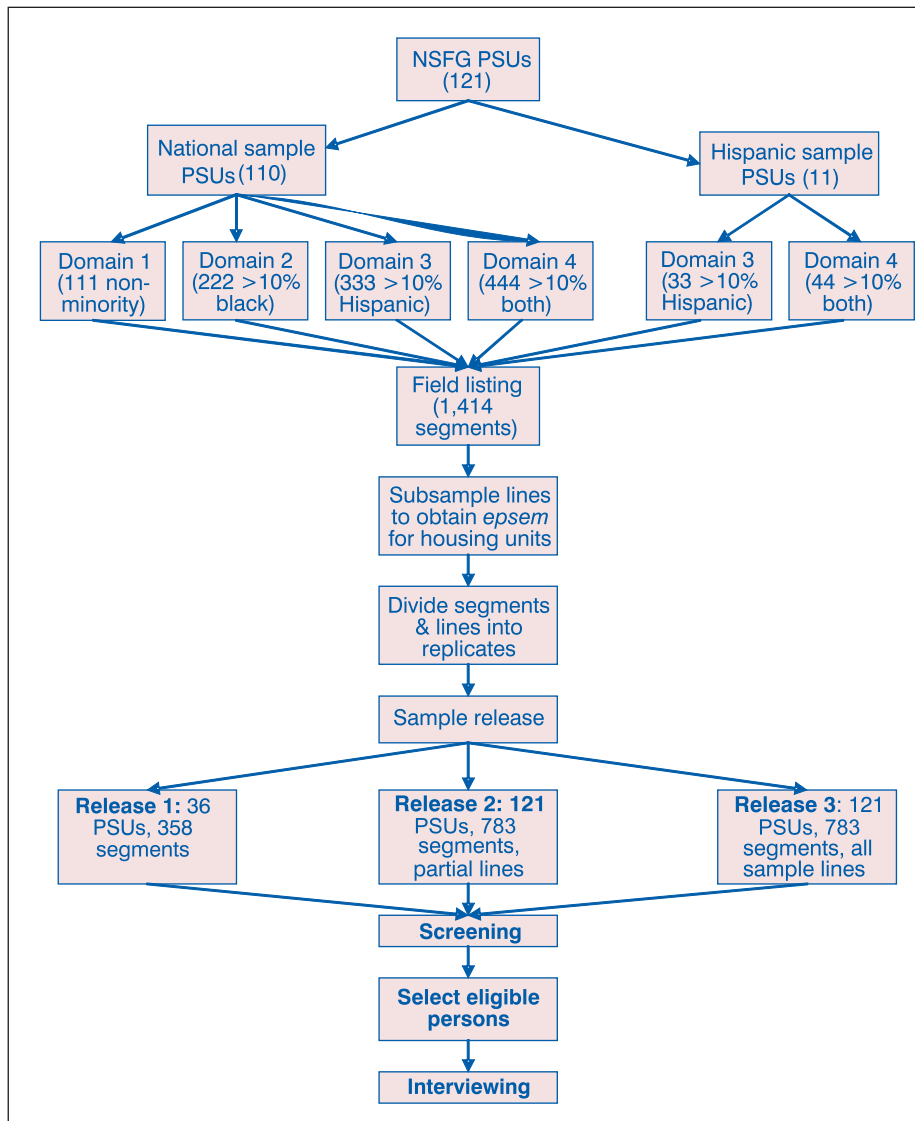
Figure 5. Selection steps in sample implementation for Cycle 6 of the National Survey of Family Growth

randomly in every sample household that contained one or more eligible persons. The within household selection used pre-assigned measures of size based on gender, age group, and race or ethnicity. Size measures were assigned to increase the chance of selection for the oversampled groups.

Persons living away from home in college or university dormitories, fraternities, and sororities were listed in their household of usual residence. If a college student were chosen in the last stage of selection, the sample person was transferred to an interviewer working in the nearest PSU to the college or university for interviewing. In some cases, interviews were obtained at the sample household when selected students returned home for summer or holidays.

Questionnaire Development

Summary of Questionnaire Development

The attempt in the pretest to provide data to meet a wide range of needs resulted in pretest questionnaires that would have yielded a very rich and useful data set, but they were too long for respondents, and too complex for interviewers. As a result, they raised the cost and reduced the response rate in the pretest and thus yielded fewer completed interviews than expected. It also appeared that the pretest questionnaires and procedures may have been too complex for some interviewers to master in the short time available for the study. Accordingly, the NSFG staff, sponsors, and contractor reached a consensus that the pretest questionnaires were too long and too complex, and worked together to shorten and simplify them, as described in the following text.

Two revised NSFG Cycle 6 questionnaires, one for females and one for males, were thus programmed into a CAPI software system called Blaise® (http://www.westat.com/blaise). The questionnaires are based on complex contingent logic, which tailors the wording of questions to the circumstances of the respondent, automatically skips the interviewer to the appropriate next question based on earlier answers, and alerts the interviewer to inconsistencies in responses. The male and female questionnaires differed in the content and sequence of questions, but also contained similar questions for some key substantive domains. This section provides more information on the content of the questionnaires and the process by which the pretest instruments were revised for the main study.

CAPI and ACASI Development for the Main Study

One of the conclusions reached after the pretest was that both the male and female questionnaire instruments were too long. In order to keep respondent burden at a reasonable level and to meet the sample size goals of the main study within the available funds, the length of the questionnaire had to be reduced. Based on detailed item-level timings and frequency information gathered during the pretest, cuts were made to the questionnaires. These cuts were strategically chosen so as to preserve the integrity of the instruments, meet the most important data needs, and reduce both their average length and the variability in their length. During this

revision process, the results from the pretest wording and order experiments were also incorporated into the questionnaire. The decisions on what changes to make to the pretest questionnaire were made in a series of meetings during 2001. Input and advice was given by NSFG staff at NCHS, funding agency representatives, and outside experts (see figure 3).

The NSFG staff at NCHS led the efforts to implement the revisions in the questionnaire specifications, aided by the NSFG staff at ISR. Once the revisions were made, the NCHS staff met with the ISR programmers to discuss the new instrument specifications. The goal of these meetings was to allow NCHS staff an opportunity to communicate the intent of the instrument, and to allow the programmers to receive immediate clarification on any parts of the specifications that were unclear. The female and male questionnaires were revised and reprogrammed on a coordinated schedule with constant discussion and refinement by e-mail and telephone between NCHS and ISR staff. While the programmers worked on one instrument, the testing staff was checking the other instrument. This schedule made efficient use of the available resources, and avoided unproductive downtime for either group.

Testing

Once the programmers completed the initial revisions, staff at ISR rigorously tested the instruments. Identified problems were fixed and the instruments were sent to NCHS for more testing. Overall, there were three major rounds of testing for each instrument (male and female) and several additional rounds to test the integration of the instrument into SurveyTrak (the ISR sample management system) and the Spanish version of the instrument.

For the most part each instrument was tested as an integrated unit, including the screening portion of the questionnaire and the sex-specific interview. The early testing was done without the ACASI sound files, but

those were added for the later rounds of testing. The first round of testing focused on the logic, flow, and edit checks in the instrument and made sure the text of the questions matched the CAPI Reference Questionnaire (CRQ) that was developed to direct the CAPI programming. Later rounds of testing checked problems identified during the earlier rounds and examined screen formatting, electronic question definitions, and ACASI sound files. Experienced field interviewers who had worked on the pretest did additional testing. Their main purpose was to duplicate an actual interview situation in order to evaluate the length and usability of the instrument. These field tests proved to be very successful, and showed that the changes made to the instruments had improved their usability and substantially reduced their length. (The inherent limit of field testing on a small scale like this, however, is that very rare problems may not be discovered, and in fact, most of the errors that were discovered after main study data collection were the result of rare scenarios.)

Spanish Translation

Given the growth of the Hispanic population into the largest minority group in the United States (22), it is critical for the NSFG to obtain accurate information about the attitudes and behaviors of Hispanics. Census 2000 showed that 49 percent of Hispanic persons over age 5 who speak Spanish do not speak English very well (23). Therefore, it was important to have a Spanish version in Cycle 6 of the NSFG. Of the total of about 2,700 Hispanic respondents, about 1,000 were interviewed in Spanish. For more information on the translation procedures used in Cycle 6, see reference 24, which is the source of the following summary.

The Spanish translation of the NSFG instruments was performed primarily by Research Support Services (RSS), a firm specializing in translation, with further input and analysis by Dr. Barbara Marin of the University of California-San Francisco and Dr. Gladys Martinez of the NSFG staff. Drs. Marin

and Martinez concentrated particularly on ensuring that the RSS translation achieved both cultural appropriateness and substantive accuracy.

One challenge was that the translation had to be sensitive to differences in vocabulary choices among major groups of Hispanics, including Mexicans, Puerto Ricans, Cubans, and others, given that the survey was to be administered in Spanish to a nationwide sample of Hispanic persons. In addition, the Spanish version had to be appropriate for less educated and newly arrived Spanish-speakers.

At RSS, translations are usually done using a traditional committee approach where three translators work simultaneously and independently, performing direct translations of the survey instruments (25). After the translations are completed, discrepancies are reconciled, and a version based on the independent translations and enriched by the group discussion is accepted. A referee, with extensive experience in survey instrument translations, chairs the reconciliation meeting. The strength in this model comes from the consensus among bilinguals that result in more accurate text than one person's translation because problems of personal idiosyncrasies and uneven skills in either language are overcome.

For the NSFG, a modified committee approach was used, in which the original translation was performed by three translators, doing one-third of the instrument each, instead of three whole independent translations (26). This approach had been used extensively by RSS and has been found to maintain the strength of the traditional committee approach, while being less expensive and less time consuming. The division of material among the translators is done in a way that avoids giving entire sections or modules to a single person. Instead, pages are sorted into three piles, one for each translator, ''in the alternating fashion used to deal cards in card games'' (27). This assures that all three translators are familiar with the different topics covered in the instruments.

The modified committee approach does not include back translation (giving a second translator the Spanish instrument and having him/her translate it back into English in order to compare the two versions of the instrument). A major drawback of the back-translation method is that even when the back translation text is identical to the original, this does not give the researcher any sense of whether respondents will be able to understand the translated version.

The committee that worked on the NSFG translations included translators who were native speakers of some of the main varieties of Spanish spoken by the Hispanic population in the United States (one from Mexico, one from Puerto Rico, and one from South America). Two of the translators were males, while the other translator and the referee were female. The committee translated the female questionnaire, the male questionnaire, and the accompanying materials (e.g., informed consent letters). The translation reconciliation sessions took a total of 21 hours. Items or terms were flagged for further research when the team was unable to reach an agreement.

After several iterations of the questionnaires, a subset of items was tested in cognitive interviews in order to identify questions that presented problems and reasons for the existence of these problems. RSS interviewed nine Spanish-speaking women and nine Spanish-speaking men. All were immigrants from eight different countries in Latin America. Their ages ranged from 16 to 41 and their age at the time of immigration also ranged widely, from age 10 to 36. Their years of schooling varied from 3 to 19.

The cognitive interviews consisted of three types of questions: questions that tried to ascertain participants' understanding of specific words for which familiarity to speakers was uncertain; questions asking about definitions; and questions about hypothetical situations in the form of vignettes. The Spanish language instruments were edited based on information learned from the cognitive interviews.

Almost all NSFG materials were translated into Spanish. Because no information about the households existed ahead of time, the advanced household letters were double-sided with an English version on one side and Spanish translation on the other side. The Life History Calendar was not translated into Spanish but there was a Spanish bookmark that could be placed on top of the English calendar.

The computer-assisted interviews were programmed using Blaise, a data collection software system. Blaise allowed interviewers to switch the language of the instrument with a single keystroke. The ACASI portion of the interview was also translated into Spanish with an audio part in Spanish. The recorded voice was from a woman of Colombian descent as previous research shows that the accent can be understood by a majority of Spanish speakers (28).

Once the Spanish language version of the questionnaire had been finalized, it had to be incorporated into the Blaise program. Using the English language program as a template, the Spanish text was copied and pasted into the program by ISR staff. Bilingual testers at NCHS and at ISR were used to make sure the Spanish instruments had been constructed properly. Specifically, they examined the instruments to make sure there were no English words showing up in the Spanish questions and that all of the "fills" worked correctly. Their review provided an additional check of the basic instrument flow and structure. Finally, NCHS tested the Spanish CAPI program when the Spanish instrument was finished.

Overview of the Questionnaires

A brief outline of the female questionnaire is shown in figure 6, and an outline of the male questionnaire is shown in figure 7. A more detailed description of the questionnaires appears in the following text.

Female Questionnaire

Female Section A: Introduction, Calendar Instructions, Demographic Characteristics, Household Roster, and Childhood Background—Many studies have shown that demographic characteristics—such as age, race, education, employment, and family background—are closely related to outcomes such as marriage, cohabitation, family size, and contraceptive use. This first section began by gathering these basic demographic characteristics about the respondent and the members of her household. The household information included the age, sex, race, and Hispanic ethnicity of each household member, and their relationship to the respondent (figure 6).

In section A, the female respondent was introduced to the Life History Calendar and asked to record several memorable events from her life on it. This calendar was used throughout the interview to help the respondent remember dates. (More information about the calendar is given later in this section.)

Next, respondents were asked about current attendance at school, highest level of education completed, and whether they had a high school diploma or GED and college degrees. Respondents were asked about whether they lived with both parents during their entire childhood, and if not, their living arrangements at age 14. Finally, questions were asked about the level of education of her mother and father (or parent figures), and the total number of children the respondent's mother had.

Female Section B: Pregnancy and Birth History, Adoption and Nonbiological Children—One of the primary purposes of the NSFG is to provide data on women's lifetime experiences with pregnancy and childbearing. This section of the survey covered pregnancies, biological children, and adoption. First, the respondent was asked about whether she was currently pregnant and how many past pregnancies she had had. For each pregnancy, the respondent was asked about the pregnancy length (i.e., gestational length) and the pregnancy outcome. If the pregnancy resulted in a live birth, she was also asked about the baby and about the delivery. For pregnancies within the last 5 years,

Section A: Background, Demographic Information

Age, Marital/Cohabitation Status, Race/Ethnicity, Household Roster, Life History Calendar Introduction, Education, Childhood/Parental Background

Section B: Pregnancy History and Adoption-Related Information

Menarche, Current Pregnancy Status, Number of Pregnancies, Detailed Pregnancy History, Age of Father of Pregnancy, Relinquishment for Adoption, Care of Non-Biological Children, Adoption Plans (current and past) and Preferences (current seekers only)

Section C: Marital and Relationship History

Marriage and Cohabitation History, Husband/Partner Characteristics, Timing of First Sexual Intercourse and Characteristics of First Partner, Reasons for Not Having Sex (among virgins), Sex Education (teens only), Number of Sexual Partners, Recent (past 12 months) Partner History

Section D: Sterilizing Operations and Impaired Fecundity

Sterilizing Operations, Desire for Reversal (for tubal ligations & vasectomies), Non-Surgical Sterility & Impaired Fecundity

Section E: Contraceptive History and Wantedness

Ever-Use of Methods, First Method Use, Periods of Non-Intercourse (last 3 years), Method Use Each Month (last 3 years), Method Use with Partners (past 12 months), Wantedness of Pregnancies, Consistency of Condom Use, Frequency of Sex (past 4 weeks)

Section F: Family Planning and Medical Services

Birth Control and Medical Services (past 12 months), Provider & Payment Information for Each Visit (more detail if clinic cited) and Whether Regular Source of Medical Care, First Birth Control Service (date and details), Ever Visited a Clinic

Section G: Birth Desires and Intentions

Wanting Another Baby, Intending Another Baby (*joint or individual as appropriate*), Number Intended

Section H: Infertility Services and Reproductive Health

Infertility Services, Vaginal Douching, Health Problems Related to Childbearing (including PID and disability), HIV Testing

Section I: More Background, Demographic Information, and Attitudes Questions

Health Insurance, Residence, Place of Birth, Rent/Own/Payment for Current Residence, Religion, Work Background and in Past Year, Current or Last Job, Child Care; Attitudes about Premarital Sex, Parenthood, Marriage, Cohabitation, Gender Roles, Condom Use

Section J: Audio CASI

General Health, Number of Pregnancies, Substance Use, Sex with Males (including Nonvoluntary Sex and STD/HIV risking behaviors), Sex with Females, Condom Use at Last Sex of Any Type, Sexual Orientation & Attraction, STDs & HIV, Family Income, Public Assistance

Figure 6. Brief outline of Female National Survey of Family Growth Questionnaire: Cycle 6 main study

Section A: Background Information

Age, Marital/Cohabitation Status, Race/Ethnicity, Household Roster, Education, Childhood/Parental background, Marriage and Cohabitation

Section B: Sex Education, Vasectomy, Infertility, Sexual Intercourse, Sexual Partners

Sex Education, (teens only), Vasectomy, and Infertility, Sexual Intercourse, Reasons for Not Having Sex (among Virgins), number of Biological Children, Number of Sexual Partners in Life and Last 12 Months, Listing of up to 3 most recent partners

Section C: Current Wife or Cohabiting Partner

Cohabitation and Marriage, Demographic Characteristics, First Sex, Sterilizing Operations and Infertility, Last Sex, Summary of Method Use in Last 12 Months, Biological Children, Current Pregnancy, Her Children, Other Children Under their Care

Section D: Recent Sexual Partner(s) (up to 3) and First Sexual Partner

Cohabitation and Marriage, Demographic Characteristics, Last Sex, First Sex, Summary of Method Use in Last 12 Months, Biological Children, Current Pregnancy, Her Children, Other Children Under their Care, First Sexual Partner

Section E: Former Wives and First Cohabiting Partner

Cohabitation and Marriage, Demographic Characteristics, Biological Children, their Children, Other Children Under Her Care

Section F: Other Biological and Adopted Children, Other Pregnancies

Additional Biological Children, Additional Adopted Children, Pregnancies that Did Not End in Live Birth

Section G: Fathering

Activities with Coresidential Children; Activities with Noncoresidential Children; monetary support of Nonresidential children

Section H: Birth Expectations

Desires and Intentions for a Future Birth

Section I: Health Conditions, Access to Health Care, and Receipt of Health Services

Usual Source of Medical Care, Health Insurance, Use of Family Planning Clinic, Disability, Health Services in Last 12 Months, Medical Help for Infertility, HIV Testing

Section J: More Background Information and Attitudes Questions

Residence, Place of Birth, Rent/Own/Payment for Current Residence, Religion, Military Service, Work Background and in Past Year, Current or Last Job (Respondent and Partner), Child Care, Attitudes about Premarital Sex, Parenthood, Marriage, Cohabitation, Gender Roles, Condom Use

Section K: Audio CASI

General Health, Significant Life Events, Substance Use, Pregnancy/Abortion, Sex with Females, Sex with Males (including STD/HIV risk behaviors), Condom Use at Last Sex of Any Type, Sexual Orientation & Attraction, STDs & HIV, Family Income, Public Assistance

Figure 7. Brief outline of Male National Survey of Family Growth Questionnaire: Cycle 6 main study

information was collected on cigarette smoking, prenatal care, and maternity leave. For children currently 18 years of age or younger, the respondent was also asked about breast-feeding when they were babies.

After pregnancy information was collected, the questions focused on adoption. The NSFG is among the few sources of nationally representative data on adoption. First, the respondent was asked about any children she had placed for adoption. Then she was asked about nonbiological children who may have lived with her under her care, and whether she adopted or became a legal guardian for any of them. Finally, adult women (18–44 years of age) were asked about their current and previous pursuit of adoption. Women currently seeking to adopt were asked about their preferences for the characteristics of the adopted child (for example, with regard to age and race). This information is important for describing the current demand for adoption in the United States.

Female Section C: Marital and Relationship History—In this section, data are provided on marriage, cohabitation, and sexual activity. These data are used to understand family formation and dissolution, as well as the patterns of sexual relationships. If the respondent was currently, or ever had been married, she was asked about those marriages and husbands. If she currently lived with a male sexual partner or had ever cohabited with male partners, she was asked about those relationships as well. For the current and first husbands or cohabiting partners, women were asked to provide their education, race and Hispanic origin, whether they had been married before, and the number of children from prior relationships.

The latter half of Section C focused on whether or not the respondent has ever had heterosexual intercourse, and if so, on recent sexual activity. (Women who have ever been pregnant, or have ever been married or cohabited, were assumed to have had heterosexual vaginal intercourse. All other women were directly asked if they have ever had sexual intercourse.) Basic information was collected about the respondent's first intercourse, her first male partner, and male sexual partners

she had had within the past 12 months. If the woman reported a recent sexual partner as "current," she was asked about his education and race and Hispanic origin. Teen respondents were asked about their experience with sex education.

Female Section D: Sterilizing Operations and Fertility Problems—In this section of the survey, questions were asked about surgical sterilizations and fertility problems. Data on surgical sterilizations are needed to measure how many women use this sterilization for contraceptive reasons. If the respondent reported that either she or her current husband or cohabiting partner had ever had a sterilization operation, she was asked the date of the operation. For each operation within the last 5 years, respondents were asked where it was done, the method of payment, and the reasons for the operation. Any respondent reporting tubal sterilization or vasectomy as her and her husband/partner's only operation was asked about the desire for reversal. The respondent was also asked if she was sterile for reasons other than surgical sterilization, if she would have any physical difficulty getting pregnant or carrying a child to term, or if her partner had any physical problems with fathering a child.

Female Section E: Contraceptive History and Pregnancy Wantedness—Contraceptive use and unintended pregnancy are key topics for the NSFG and they are the focus of this section. Extensive data were gathered regarding the birth control methods used by the respondent, including all types of contraceptives ever used, and the first method she ever used. She is then asked to use the Life History Calendar to report contraceptive methods she used each month during the past 3 years.

The reason for asking this detailed information on intercourse and contraceptive use is so researchers can analyze how well the methods work in preventing pregnancies (29). The NSFG also provides important national data on estimates of unintended pregnancies among U.S. women. Respondents are asked to report whether or not each of their pregnancies was wanted (30). They are also asked about their current

pregnancy intentions, current methods of birth control, and source of that birth control method.

The section concluded with a few specific questions about birth control pills (the most commonly used method in the United States apart from sterilization). Women who currently use or recently used the pill were asked to report their pill brand, with the help of a Pill Chart showing virtually all the pill packs on the market in the United States in 2002. Researchers use these data to relate hormonal dosage of birth control pills to use-effectiveness and health consequences for women.

Female Section F: Family Planning and Medical Services—Data on the use of family planning and medical services are essential to an understanding of what services are needed by the population. In this section, the respondent was asked about use of birth control and reproductive-related medical services during the last 12 months, including service providers and methods of payment (31,32).

Female Section G: Birth Desires and Intentions—Questions in this section focused upon the respondent's desire to have children in the future and her intentions to actually have children. This information is used to track trends over time in expected family sizes.

Female Section H: Infertility Services and Reproductive Health—The NSFG is the only nationally representative source of information on the use of medical services for infertility in the United States. In this section, the respondent was first asked whether she had ever received medical help to aid her in getting pregnant. After a positive answer, she was asked about the specific types of medical help she or her partner(s) received. Following this series, the respondent was asked about any medical help she may have ever received to prevent miscarriages. Taken together, these series of questions help to measure the use of various treatments and services for infertility. Women were also asked if they were told that they or their partners had specific infertility problems, such as ovulatory problems, blocked fallopian tubes, or semen problems.

Questions about other behaviors and conditions that might affect childbearing and health then followed, including questions on vaginal douching, pelvic inflammatory disease (PID), diabetes, and physical disability. Given that HIV/AIDS is a critical health concern, the section concluded with questions about HIV testing and counseling. Recently pregnant women were also asked about their knowledge about retroviral treatment to prevent perinatal HIV transmission.

Female Section I: Insurance, Residence, Work Experience, and Attitudes—First the respondent was asked about her health insurance coverage over the past year. This was followed by questions about the respondent's current residence, her residence at the time of the 2000 Census, and whether she was born outside the U.S. These questions are necessary because neighborhood characteristics can have important influences on the sexual, contraceptive and marital behavior of the men and women living there. Several questions about the respondent's religious affiliation and attendance were also asked.

Research has documented effects of employment on childbearing and marriage. Section I includes a short series of questions about the respondent's work experience and the current or most recent job of the respondent's husband or cohabiting partner, if she had one. Women living with any children under 13 years were asked about child care arrangements they may have used during the last 4 weeks.

This section concluded with a set of questions about the respondent's attitudes and opinions about marriage, gender roles and parenthood, sex, and condom use.

Female Section J: ACASI (Audio Computer-Assisted Self-Interview)—This section gave the respondent an opportunity to answer a series of sensitive questions privately using the computer and headphones. The first few questions served as practice to familiarize the respondent with the mechanics of the ACASI section. Once the laptop had been turned over to the

respondent, general health questions were asked, followed by questions about cigarette, alcohol, and drug use; numbers of pregnancies; sexual behavior with males and females; non-voluntary sexual intercourse with males (asked only of adult respondents 18–44); STD/HIV risk behavior with males; STD experience in the last 12 months; and income and receipt of public assistance.

Male Questionnaire

Male Section A: Demographic Characteristics; Household Roster; Childhood Background; Numbers of Marriages and Cohabitations—This section was virtually identical to female Section A, except that the male questionnaire did not use a Life History Calendar, so this section does not introduce one. The other key difference in this initial section was that men were asked how many times they had been married or cohabited. The information was needed up front in the male questionnaire because of how subsequent sections were organized.

Male Section B: Sex Communication and Sexual Experience—This section obtained information on male sterilizing operations (most commonly a vasectomy), male infertility, and the respondent's sexual experience. For adult respondents (18 or older), this section began with the questions about sterilization operations and infertility (figure 7). If the respondent reported a sterilizing operation within the last 5 years, he was asked about the place where the operation was performed and how it was paid for.

This section was also designed to measure the population that is sexually active (and at risk of having children) and to estimate changes in sexual behavior over time. First, respondents were asked about prior sexual experiences with female partners. Each respondent who had had intercourse was asked about any biological children he may have fathered, and the total number of female sexual partners in his lifetime and in the past 12 months.

Sections C, D, and E obtained essentially similar data to those obtained in female Sections B, C, D, and E, but

the male and female questionnaires were structured somewhat differently to ask about the current wife or cohabiting partner in Section C, recent partners in Section D, and former wives and the first cohabiting partner in Section E.

Male Section C: Current Wife or Cohabiting Partner—Currently married or cohabiting men provided information about their current wife or cohabiting female partner in this section: dates of marriage and cohabitation, demographic characteristics of the wife or cohabiting partner, her experience with surgical sterilization and fertility problems, and their contraceptive use. The respondent was then asked about biological and adopted children he has had with his current wife or cohabiting partner, as well as any children she may have had from previous relationships.

Male Section D: Recent Sexual Partners and First Sexual Partner—The NSFG collects information in Section D on the characteristics of up to three recent female sexual partners in the 12 months preceding the interview. Even if the respondent reported no sexual intercourse in the last 12 months, he was asked about his most recent sexual partner. Partner characteristics are important in assessing the changing patterns of condom use—trends central to the prevention of teen pregnancy, unintended pregnancy, STDs, and HIV. Data on his partners included their demographic characteristics, dates of first and last sex, contraceptive methods used, and pregnancy outcomes.

Since one purpose of the survey is to learn more about fatherhood, men who had married or lived with any of these recent female sex partners were asked about any biological or adopted children he had with that partner, as well as any children she may have had from previous relationships. These questions about children are similar to the questions described above under Section C.

This section ended with questions about the respondent's first sexual intercourse, including his age at that time, some basic information about that partner and contraceptive use.

Male Section E: Former Wives and the First Premarital Cohabiting Partner—Another goal of NSFG is to describe men's marital and cohabitation experiences. This section collected information about former wives and the first female cohabiting partner. Questions about children with former wives or cohabiting partners were similar to those described under Section C above.

Male Section F: Other Biological Children, Other Adopted Children, Other Pregnancies—This section completed the man's fertility history by collecting information on any other biological children he had fathered with women not previously discussed in the interview. The questions about these children were similar to the ones asked in Sections C–E. However, the respondent was also asked a few questions about the mother of each child since she was not previously discussed (for example, her age at time of the birth). Men 18–44 were also asked about any children they may have adopted on their own.

Male Section G: Fathering—This section provided important information on the respondent's fathering behavior, both with his children (biological and adopted) who live with him and his children who live elsewhere. The specific activities or behaviors asked about were based on the age of the child. Some questions were asked for children under 5 years (for example, *"How often do you bathe, diaper, or dress* (him/her/them)?'' and some were asked for children 5–18 years old (for example, *"help with homework"*). We also asked for the respondent's opinion of ''how good a job'' he was doing as a father. For his biological and adopted children who live elsewhere, he was asked about financial support of these children.

Male Section H: Desires and Intentions for Future Children— Questions in this section asked the respondent about his desire and intention to have children in the future and the number of children he would like to have. This information will be used to help track trends over time in expected family sizes.

Male Section I: Health Conditions and Health Services—Information from this section can be used to inform policy makers about health care use and needs among men. The respondent was first asked about his usual source of health care and his health insurance coverage over the past 12 months. Men were then asked about specific services they may have received in the 12 months preceding the survey, including a routine physical exam, screening for testicular cancer, counseling about contraception or sterilization, and testing or treatment for STDs, including HIV and AIDS. If respondents under age 25 reported receiving any of these services in the last 12 months, they were asked further questions about the providers of these services and how they were paid for.

This section also collected data on use of infertility services. The female NSFG has long been the only nationally representative source of information on the use of medical services for infertility in the United States. The male NSFG provided an opportunity to learn more about infertility services from the male perspective. If the respondent had ever received this kind of help, followup questions were asked concerning the specific services he received as well as any diagnoses he may have received.

The last series of questions in this section were on HIV testing and knowledge, a critical health concern in the United States. These questions were essentially the same as the series asked of females in Female Section H.

Male Section J: Residence, Work Experience, and Attitudes—This section gathered important background information about the respondent and his wife or cohabiting partner. This information was in large measure similar to the information collected in Female Section I.

Section J included a short series of questions about the respondent's work experience and military service. This section concluded with a set of questions about his attitudes toward relationships, sex, condom use, gender roles, and parenthood similar to the attitude questions asked of women.

Male Section K: ACASI (Audio Computer-Assisted Self-Interview)—This section gave the respondent an opportunity to answer a series of sensitive questions privately using the computer and headphones. The questions were very similar to those asked of women in audio CASI. Notable differences included: men were asked whether they had recently been in a homeless shelter or jail (as a measure of turnover from those populations into the household population), and men were asked more questions related to same-sex sexual behavior than women were, due to the greater risks for HIV and STD transmission.

Materials Used to Administer the Interview

Both the male and female questionnaires used a ''Showcard Booklet'' and on-line help with questionnaire specifications.

● Both male and female respondents were given a ''Showcard Booklet'' listing response options on some questions.
● Question by Question specifications, or "Q by Q's," were available for most questions in the interviewer-administered portions of the male and female questionnaires. The interviewer accessed the specifications by pressing the "help" key (F1). These entries defined key terms, and explained the intent of each question. They provide a more modern and accessible alternative to paper Q by Q's.

For the female interview, two additional aids were used.

● Female respondents had a ''Pill Chart,'' which displayed pictures of various brands of oral contraceptive pills. This chart was shown to each current or recent user of oral contraceptive pills to help her identify the brand and type she used.
● Female respondents also used a 31″ × 12″ paper Life History Calendar to help with the recall of dates of events asked in the interview. A 2.25″ × 12″ laminated sliding column was also provided to help the respondents keep their place on the calendar. Early in the interview, the interviewer described

the purpose of the calendar and helped the respondent fill in some important events in her life. During the interview, the female respondents were reminded to refer to the calendar to help them recall dates and then to mark those dates on the calendar.

Interviewer Materials

Manuals

The manuals describing the details of the NSFG data collection procedures were used by field staff in training and as a reference during field work. The 346-page *National Survey of Family Growth Cycle 6 Interviewer Project Manual* was sent to each interviewer trainee about 1 week before training, along with home-study questions, which the trainees were asked to complete and bring to training. The same manual, along with a 113-page *National Survey of Family Growth Cycle 6 Team Leader Project Manual* and a set of home-study questions, were sent to the Team Leader trainees about 1 week before their training.

The *Interviewer Project Manual* covered the NSFG project design; background of the NSFG project; computer equipment and software used on the survey; procedures for contacting sampled households, obtaining cooperation, screening the sampled household, and obtaining consent for the survey; an overview of the questionnaire and how to administer it; quality control measures; and how to report interviewer time and expenses. Along with the manual, the interviewers received a booklet explaining how to use SurveyTrak—the sample management software used for this survey.

The *Team Leader Project Manual* covered the general role and responsibilities of the team leader, including monitoring and managing production and costs, quality control measures, and addressing respondent concerns and reluctance. The team leaders also received two booklets explaining how to use the production monitoring software used in the survey.

Materials Used for Locating Households and Contacting Respondents

Before the interviewers attempted to contact the sample households, advance letters were mailed to sample households (see Appendix III). This mailing was done from the Ann Arbor office at the time the sample was initially released to the field. The letter introduced the study to the household and informed them that an interviewer would be contacting them soon.

The Segment Folder listed the addresses for the group of neighboring sample housing units selected from a sample "segment." A Segment Folder was created for each segment that contained sample households, and included a map of the general area where the segment was located, a detailed map of the segment itself, and a list of all the households in the segment.

After locating a sample household, interviewers attempted to make contact and conduct a screening interview. The screening interview determined whether any of the household members were eligible to participate in the main interview. A screener introduction script (see Appendix IV) was developed to help the interviewer introduce the study to the household members and begin the screening. Interviewers were provided with copies of the household advance letter to show to householders in case they did not receive or did not remember receiving the advance letter.

Materials Used for Obtaining Participation in the Survey

Several items were used to help obtain the participation of members of the sample household:

- The Household Advance Letter (see Appendix III) was used to make initial contact with residents and to help gain cooperation.
- Once a respondent had been selected to complete the survey, a Respondent Advance Letter (see Appendix III) was used. It resembled the household advance letter, but emphasized the issues of

the main interview rather than the screening interview.
- Another set of letters was developed to address various concerns raised by the household members or respondents on the initial contacts made by the interviewer. The concerns addressed in the letters were: being too busy, feeling the interview was too personal, not being interested in the survey, and being generally reluctant to participate. There were also letters targeted at household members who were difficult to find at home, parents or guardians who were reluctant to consent to their child's participation in the survey, and managers of locked buildings. All of these letters were sent by the team leaders after consulting with interviewers about the circumstances of the reluctance.

Interviewers were provided with several other types of materials:

- A picture identification badge
- Contact scripts to help them introduce themselves and the survey (see Appendix IV)
- A Letter of Authorization verifying the interviewer's position as an NSFG interviewer (see Appendix V)
- The NSFG Question and Answer Brochure, mailed to sample households with the household advance letter (see Appendix VI)
- A Confidentiality Brochure developed for interviewers to use at the doorstep if a household member or respondent was concerned about confidentiality (see Appendix VII)
- The NSFG Family Facts sheet to demonstrate how the NSFG data were used and reported (see Appendix VIII).

Computer Hardware and Related Supplies

The hardware used in the NSFG Main Study consisted of laptop computers for the field interviewers and supervisors. The laptop computers were IBM with a track pointer, an internal mouse, an external CD-Rom drive, and two USB (Universal Serial Bus) ports. The related computer supplies included

an AC adaptor, an extension cord, an AC car adaptor, headphones for use during the ACASI portion of the questionnaire, a telephone cord, a telephone cord adaptor, and a three-prong power cord adaptor. The staff was provided with a laptop case in which to carry all of these items. The supervisors were also given a printer and an external mouse.

Interviewer Training

Interviewer Demographic Characteristics

All interviewers for the NSFG Cycle 6 were female. About 14 percent of the 262 interviewers for the NSFG Main Study were of Hispanic origin, 49 percent were non-Hispanic white, 26 percent were non-Hispanic black, and 6 percent were of some other race. Among all interviewers, 23 percent spoke Spanish, 61 percent had previous experience as interviewers and one-third of those were experienced in both face-to-face and telephone interviewing. Only 3 percent of the interviewers had worked on a previous cycle of the NSFG. Among the interviewers, 8 percent had worked in a study with sensitive content such as sexual activity, drug use, or criminal content. Almost all interviewers had used a computer before (98 percent); and 26 percent had a bachelor's degree. About three out of four interviewers (73 percent) considered interviewing their main occupation.

General Training Program

The first interviewer training session for the main study was held in March 2002, and 120 field interviewers were trained. In June 2002, 118 additional field interviewers were trained. Most of those trained in June were brought on to staff new sample areas, but some staff was trained to make up for attrition among the March trainees. In September, another 52 interviewers were trained, in part to replace terminated interviewers. In total, 290 field interviewers were trained for this project; about 270 of these

completed at least one screener interview; others did not meet training criteria or otherwise ended their employment before completing a screener interview. An outline of the training program appears in Appendix X. Each session was held in a hotel large enough to house and train all the interviewers in that session in a single place. The locations were chosen based on cost of the space, ease and cost of travel to the location, and the suitability of the meeting space for interviewer training.

The training program had four parts: pre-classroom home study, general interviewer training, NSFG project-specific training, and Spanish bilingual training. In addition, study halls were provided on several nights during training for interviewers who were having difficulty or had additional questions.

Preclassroom home study training was a self-study of the *Interviewer Project Manual* with a written exercise. Interviewers who had not interviewed before for the University of Michigan or at all were asked to attend a 1-day general interviewing techniques session where they were taught the fundamentals of good interviewing technique and how to complete the forms used to record time and valid expenses incurred on the job.

For NSFG project-specific classroom training, the interviewers were divided up into classrooms with 20–25 trainees. Each classroom had a lead trainer, an assistant trainer, a data display operator, and a runner. The roles of lead and assistant trainer were usually filled by central office ISR staff and regional field managers. The data display operators and runners were usually team leaders. The lead and assistant trainers both gave lectures on data collection procedures and administrative tasks, and they both led exercises and hands-on practice. Part of the training on administering the questionnaire involved four ''round-robin'' interviews, during which each classroom divided into two smaller groups. The lead and assistant trainer each led one of the smaller groups.

The first day of project-specific training was devoted to learning how to

use the computer and the related software, contacting households, and screening the households. The information presented during this day provided the foundation for the interviewers to learn how to administer the questionnaires themselves.

The next 4 days were devoted to learning about the questionnaires and how to administer them. The female and the male questionnaires and the tools used to complete them were introduced to the interviewers through lecture sessions. Then each classroom divided into two smaller groups of 10–13 for ''roundrobin'' interviews. There were two ''roundrobin'' interviews for the female questionnaire, and two ''roundrobin'' interviews for the male questionnaire. Each interview was progressively longer and more complex. Trainees took turns asking questions, the lead or assistant trainer gave the responses from a prepared script, and the trainees entered the responses into their computer.

Interspersed with these round-robin interviews were supplemental sessions on various administrative procedures and additional practice with the interview-related software. NSFG staff from NCHS and senior ISR staff were present at each training session to observe the quality of the training and to answer questions from the trainers and trainees. There was also an opportunity for interviewers to ask ISR and NCHS staff substantive questions about the survey.

Refusal Aversion Training

During the morning of the last day of training, interviewers participated in a special 4-hour workshop on how to address concerns that potential respondents may have about the study. The basic premise of the training, based on years of scientific research on survey response (33), is that potential respondents often have many legitimate questions about someone who comes to their door asking for something. Those might include questions such as:

- Who are you and what do you want from me?

- Are you selling something?
- Do you want money from me?

Once it is understood that the interviewer is there for a study, respondents may have questions such as,

- How was I chosen? Why can't you interview my neighbor?
- What is the study about? Who is sponsoring it?
- Why is it important that I participate in it?
- Do I have to do it right now?

The purpose of the workshop was to review the most common kinds of questions and concerns that respondents have about surveys, learn to recognize them, and learn to give an appropriate response that answers their question. Interviewers received a "refusal aversion" workbook in advance of the training program. The training course outlined the steps to encourage respondent participation including preparing for the interaction, diagnosing the main concerns of the persons, selecting an appropriate response and quickly delivering relevant information.

The training program had several exercises that progressively built the interviewer's skills to meet the overall objective—successfully identifying the reason for refusal and addressing the concern quickly and accurately. There were written exercises, as well as extensive role plays that progressed from simple to complex but each taught them to identify the refusal concern and an appropriate response, and quickly deliver it. There were several exercises as part of the training:

- **Exercise #1—***Case preparation triad exercise*
  The trainees were placed in groups of three, and received a refusal scenario. The exercise was a role-play where one trainee was the respondent and the other two trainees were interviewers making contact with the householder. The role-play included a case summary note that reflected the interaction and the "second" interviewer attempted to convert the refusal. The trio debriefed together to give feedback on the techniques used and

how well the case notes prepared the second interviewer for the followup contact.
- **Exercise #2—***Written exercise #1*
  After the trainees had been trained on the study's common reasons for refusal, a written exercise was completed. In this exercise the trainee had to identify the concerns for 20 separate statements that a respondent might say. (The maximum score was 20.)
- **Exercise #3—***Flashcard rounds*
  Progressively more difficult exercises were introduced. These included diagnosing the respondent's concern, identifying the appropriate reply, and delivering the response.
- **Exercise #4—***Written exercise #2*
  In this "final" written examination, 11 respondent statements were read aloud to the interviewers, and they had to record the diagnoses and then were given a limited amount of time to record a response that they would give orally to the respondent. (A perfect score was 11.)

Classes were released early in the afternoon so interviewers could prepare for the certification interviews scheduled for that evening and the next day. A certification interview attempted to duplicate many essential aspects of a real interview situation. A trainer, following a prepared script, played the part of a respondent for a trainee. He or she also evaluated the trainee on the various aspects of conducting an interview. At the end of the certification, the trainer gave the trainee verbal feedback and prepared written feedback for the ISR staff.

While other staff was completing certification interviews, interviewers who were fluent in Spanish were asked to attend the bilingual training session. During this 5-hour session, the interviewers reviewed and practiced using the Spanish version of the CAPI questionnaires. The bilingual training was conducted by training staff that were fluent in both Spanish and English. Once these trainees returned home, they completed a certification interview in Spanish over the telephone.

Data Collection

Summary of Data Collection

The field organization for Cycle 6 consisted of 32 work teams, each led by a team leader. These work teams were organized into three regions and were directed by regional field managers who lived in the regions. The field operations coordinator supervised the regional field managers and facilitated communication among them and between them and the project staff at the contractor's headquarters.

Team Leaders

The team leader's primary responsibility was to guide and support the interviewers and help them meet the production goals established for the NSFG. To accomplish these goals, the team leaders' specific duties included: monitoring daily production, effort, and costs for each of the interviewers on their team; helping interviewers develop effective techniques for gaining cooperation from reluctant respondents; managing the distribution of the workload for their team; evaluating interviewers to ensure they were following study protocols; and providing feedback to interviewers about the quality of their work.

Team leaders held weekly conference calls with their teams to discuss progress on the sample and problems the interviewers were facing, to review protocols and procedures, and to disseminate information from ISR about the study. As needed, the team leaders also conducted individual calls with each interviewer to review her sample and discuss strategies to gain cooperation from the sampled households or persons. Finally, the team leaders were always available by telephone or email to provide advice or resolve issues the interviewers were facing.

Sample management software

The field staff was supported by an electronic sample management system. This system tracked the status of each

sample case from its release to the field through its completion as a complete interview or other final result code. The system consisted of three software components, which ISR calls SurveyTrak, WebTrak, and TLTrak.

SurveyTrak resided on the interviewer's laptop, and organized her sample assignment. For each sample household, SurveyTrak stored and displayed data on:

- The day, time, and outcome of each call attempt
- Reminder notes to the interviewer about the call attempt
- Appointment day and time
- Mode of contact (face-to-face or telephone)
- Any questions asked by a household member
- Any comments made by a household member
- Sample segment information relevant to processing the case

The software was interactive, so the interviewer could update it to record her contact attempts at the sampled households and to make general notes about the cases. Active cases could be sorted by intermediate disposition codes, so the interviewer could organize her workdays. Once a day, the interviewer performed an electronic communication to send her most recent work to the ISR central computer.

The data collection was conducted in phases, in order to base the final design features on real experience regarding the percentage of sample housing units containing one or more persons 15–44 years of age; the percentage of sample housing units containing teenagers, black, and Hispanic persons (oversampled groups in the survey); the amount of interviewer time required to contact and gain the cooperation of sample persons; and the response rates achievable among females and males. The detailed description of this phased design is given in the section on "Responsive Design Features of Cycle 6."

The following sections give more details on the main study data collection procedures.

Field Organization

The sample areas were grouped into work teams using several criteria. The goal was to have manageably sized work teams made up of interviewers working in similar areas. Many of the major metropolitan areas had a team devoted exclusively to working in that city. New York and Los Angeles—two cities with traditionally low response rates, high costs, and large interviewer turnover—were each worked by two teams. The other teams in the study were also organized so that all of the team members were working a sample with similar characteristics—such as rural, suburban, or urban—maximizing the interviewers' ability to help each other strategize about difficult cases. Usually areas were grouped by geographic proximity, but sometimes demographically similar areas were put together even if they were not near each other geographically.

Interviewer Assignments

Regional Field Managers and centralized project staff worked together to assign sample cases to interviewers. The sample was assigned based on the location of the sample households to each other and to the interviewer's home. Geocoding techniques were used as much as possible in making the assignments.

Recruitment Protocol

A "recruitment protocol" is a set of procedures specified by the survey designer for interviewers and other data collection staff to use to contact, obtain cooperation, and conduct interviews with sample households. The procedures of the recruitment protocol were reviewed and approved by Institutional Review Boards at both NCHS and the University of Michigan.

The key steps in the recruitment protocol were:

1. Before contacting households in person, the contractor sent an advance letter and pamphlet to all eligible households (see Appendix III). These explained who was sponsoring the survey, who was conducting it, why it was being done, and that it was voluntary and confidential. For the main study, Spanish versions of the questionnaires, the advance letter, and other introductory materials were prepared (see Appendix III).

2. Because it was often difficult to find anyone at home, interviewers often failed to contact anyone in the sample household on the first visit and would return many times until the household was contacted. If contact was successfully made but the contacted person had no time to complete the screener, or wanted to think about it or discuss it with other household members, the interviewer would come back at a later time.

3. The interviewer conducted a brief household screening interview to determine who, if anyone, would be the selected respondent. If there were no age-eligible person (15–44 years of age) living in the household, the contact with the household was over (see Appendix IV).

4. When a field interviewer contacted a sample household, she introduced herself, displayed her identification badge, perhaps showed the authorization letter (see Appendix V) and explained the purpose of the study, referring to the advance letter that the household should have already received (see Appendix IV for interviewer scripts). Many times the interviewer found it helpful to provide sample households with the Question and Answer brochure (see Appendix VI) to answer any questions the household might have about the study.

5. When a person 15–17 years of age was selected for the sample, signed parental consent was obtained first. A parent letter and consent form (see Appendix III) were used to explain the survey to the minor's mother, father, or guardian, and ask for his or her written consent. Then the minor was asked for his or her consent. If the minor respondent did not sign the

Minor's Assent Form, the case was treated as a refusal. Emancipated minors, 15–17 year olds who were married or cohabiting and living away from their parents, were rare in a sample of this size, but when encountered, they were treated as adults. This has been NSFG practice since Cycle 1 in 1973.

6. If the respondent was 18 years of age or older, the interviewer gave the respondent an Adult Consent Form, which explained the survey and requested signed consent (see Appendix III). If the respondent agreed to do the survey but refused to sign the form, the interviewer could offer to begin the interview, and ask for a signature at the end of the interview. If the respondent again refused to sign, the interviewer was permitted to sign noting that the respondent agreed to the interview but did not want to sign the form.

7. The interviewer gave the respondent $40 in cash; the respondent signed a receipt for the $40.

8. The interview was conducted, with the interviewer using a laptop computer to read the questions and key in the answers.

9. Finally, at the end of the interviewer-administered interview, the interviewer gave the respondent a pair of headphones and the notebook computer, and showed the respondent how to make simple entries on the computer. The respondent then completed a 10 to 20 minute audio CASI. The interviewer could not see or hear what questions the respondent was being asked over the headphones, could not see or hear the respondent's answers, and could not back up later and see the answers. Moreover, no one in the household could hear or see either the questions or the answers.

10. At the end of the audio CASI section, the interviewer turned off and locked the computer, thanked the respondent, and left the housing unit.

Responsive Design Features of Cycle 6 to Control Costs and Errors

Summary of Responsive Design Features

In previous cycles of the NSFG, as in many other surveys, the survey managers attempted to keep track of interviewer effort and field costs using a paper and pencil system to record each interviewer's hours worked and other expenses. The problem with these systems was that it was often 3–6 weeks before their results were available. As a result, data on survey costs lagged far behind data on completed cases and response rates, and there was little or no information that was specific enough to provide advice to interviewers that was based on detailed knowledge of their work. As a result, the NSFG Cycle 6 contract required that the contractor have a system that could provide more timely data on field costs and response rates.

Surveys with high response rate goals and limited budgets, such as Cycle 6 of the NSFG, need a way to stay informed on how much interviewer labor and money are being spent on data collection; what areas and interviewers are having good results and poor results; and what types of nonresponse are most prevalent in each area. ISR developed a system (previously described) called SurveyTrak, to provide that information each working day.

These were especially critical questions in the NSFG Cycle 6 Main Study in 2002, because the NSFG's budget for data collection was limited, target sample sizes were set for very small subgroups, cooperation rates in other U.S. surveys had been falling, and the pretest interviews had been significantly longer than was feasible for the main study. To prevent cost over-runs in main study fieldwork, it was necessary to determine as early as possible how well the main study interviewers were performing the new tasks they had been trained to do; how many hours of effort would be required to obtain interviews, whether main study interviewers would work as many hours

as they were expected to work; and what the response rates would be. Once this information was obtained, the project staff would know how many more interviewers should be trained for the main study, and how many completed interviews they could be expected to produce within our budget.

Phase 1 of fieldwork—In March 2002, an initial group of interviewers completed training and started fieldwork in a subsample of areas. The data obtained from this group through the SurveyTrak system showed that the procedures were working well, and made it possible to make better estimates of interviewer production, set callback rules, and predict survey costs for the rest of the main study.

Phase 2 of fieldwork—Further steps were taken to manage fieldwork actively in Phase 2 of fieldwork—June 2002 through January 2003. The SurveyTrak system required interviewers to record simple characteristics of neighborhoods (sample segments), sampled households, and things that respondents said during attempts to contact them for an interview. The contractor collected and analyzed these data during fieldwork, to provide data to allocate interviewer labor during the study.

Phase 3 of fieldwork—The data from the SurveyTrak system were used throughout fieldwork, but they were used most effectively in the last month of fieldwork—February 2003—when approval was obtained to use somewhat larger incentives to encourage participation. The SurveyTrak data were used to sort the remaining cases into segments (neighborhoods) that had the largest number of people who were most likely to complete the interview. Resources were focused on those areas, and a large number of interviews were obtained in the last month of interviewing, while simultaneously improving the balance of the NSFG data set across key age groups.

The following section describes how this process worked in the main study. This section of the report presents material that is more technical in nature than the preceding sections. In contrast to previous cycles of the NSFG, large amounts of field administrative data were updated daily, made available to

the project managers, and used for statistical modeling and statistical process control analysis. These data permitted project staff to manage interviewers' work during the field period. The procedures to produce and use these data to manage fieldwork are called ''responsive design'' features and are designed to yield higher response rates in key demographic groups in the sample than would have occurred without it. Readers who do not need to understand these more technical aspects of survey management may want to skip or skim the following section.

Sampling and Field Uncertainties in the National Survey of Family Growth Cycle 6

In Cycle 6, there were target interview counts for 18 groups—males and females by three categories of age by three categories of race/ethnicity. Screening interviews with sample households collected household roster data in order to identify whether any persons 15–44 years of age lived in the household. In age-eligible households, one and only one respondent was selected for a ''main'' interview. Female main interviews required about 85 minutes to complete; male interviews, 60 minutes. The targeted response rate for females was 80 percent; for males, 75 percent.

There were several uncertainties about these goals at the time of the design of the survey:

- Eligibility rates for the 18 sub-populations could not be controlled fully and thus were a random variable at the time of the data collection. Census data, Current Population Survey estimates, and data from the National Health Interview Survey (NHIS) were used to set the design; but the utility of these are diminished because of the instability of specific data by age, sex, and race on very small area units, and the inevitable changes in population distributions between the time the data were collected and the time the NSFG was done.
- The total length of the questionnaires was difficult to

estimate before testing them in the field, because the question sets were highly dependent on characteristics of the respondent that were unknown at the time of the selection. The pretest questionnaires were much longer than the target lengths; the revised questionnaires had been timed only on a small number of simulated respondents. They had not been subjected to a large field test under real conditions.
- The level of effort required to make contact with sample households was not precisely known, given growing difficulties with gated subdivisions and locked multiunit structures. Thus, there was concern about variation in contact effort needed across the sample areas.
- The level of effort required to gain the cooperation of contacted households was not known precisely. There was evidence of reduced cooperation with surveys over time and evidence of increased interviewer hours required to obtain an interview in other repeated cross-sectional surveys, and in the NSFG Pretest. Given this evidence, there was concern about how much effort would be required to achieve the response rate targets.
- The level of effort needed to achieve stable estimates of key statistics, given the design, was unknown. The most common field protocol for a household survey like the NSFG introduces to the respondent a set of stimuli that attempt to encourage the respondent to cooperate early in the protocol. These might include an advance letter, an announcement of an incentive, and descriptive material about the survey, among others. After those stimuli are delivered, in most surveys, the protocol is merely extended through repeated contacts with the respondent until a final resolution is obtained. Such designs often reach a point where estimates are stabilized, given the fixed design. That is, the nonresponse and measurement error properties of the design are relatively fully elaborated and extending the design adds cases but changes estimates only slightly (34).

These uncertainties produce a lack of control of the cost of the data collection. For example:

1. If the eligibility rate of an age by sex by race and ethnicity group is misestimated, larger sample sizes would be required to achieve a given target sample size for the group.
2. If the length of the questionnaire was underestimated, the number of interviewer hours required to complete an interview would be higher than expected.
3. If the value of key estimates from the survey did not change in important ways after a given level of effort (e.g., 10 calls per household), it would make sense to stop spending project funds on further calls, and spend those funds somewhere else.

Responsive design makes issues such as these empirical ones during a data collection, and allows project managers to make choices based on the data.

Given these uncertainties and a fixed budget, the only way to control data quality was to allow the design to change based on production experience. The NSFG Cycle 6 used a variety of ''responsive design'' features to control costs and errors in the resulting estimates. By way of definition, responsive survey designs:

1. Pre-identify a set of design features potentially affecting costs and errors of survey statistics
2. Identify a set of indicators of the cost and error properties of those features
3. Monitor those indicators in the initial phases of data collection
4. Alter the active features of the survey in subsequent phases based on cost/error tradeoff decision rules
5. Combine data from the separate design phases into a single estimator

While some of the features of responsive design have been used in other surveys (e.g., Cycle 4 used a sample of nonrespondents at the end of the data collection in an attempt to reduce nonresponse error), Cycle 6 was one of the first surveys to use real-time administrative data to direct the field

NOTES: "Pr(sel)" means probability of selection. "Informant Behavior" is what the person who answered the door did or said.
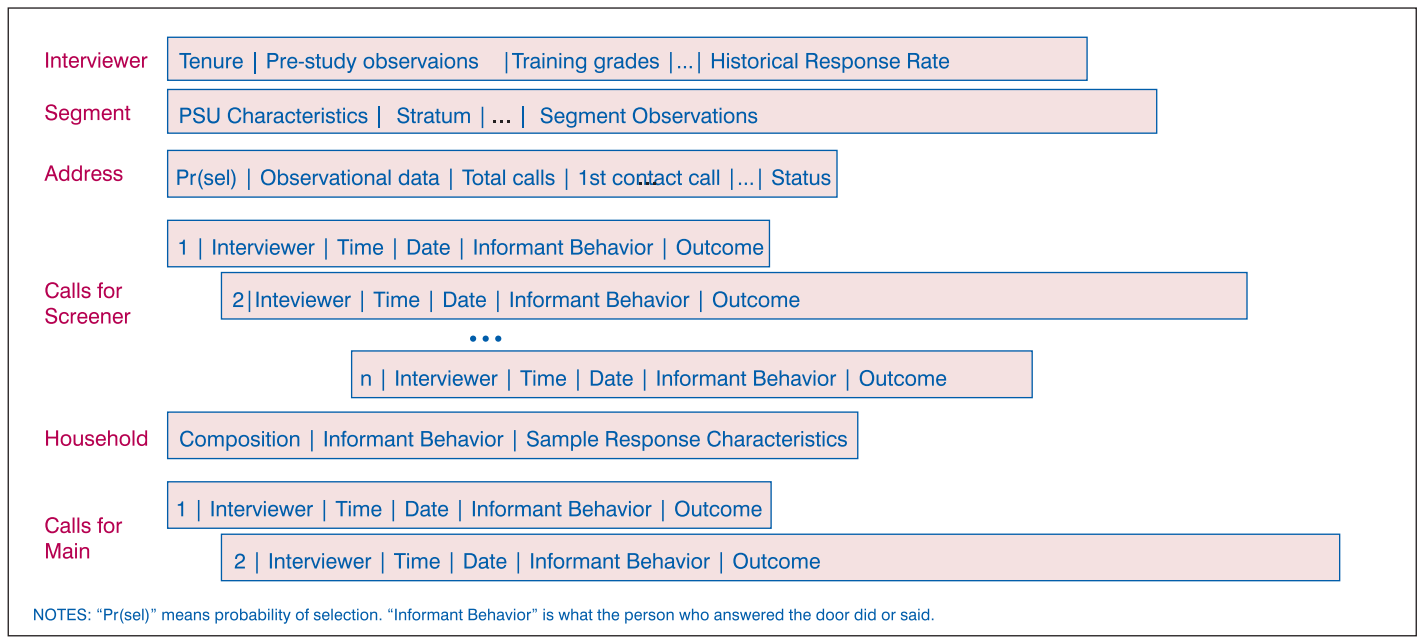
Figure 8. Data structure for National Survey of Family Growth paradata or process data

efforts (35). As noted above, the goals of this field management effort were to manage costs and effort as efficiently as possible to attain response rate goals and to prevent cost increases.

Observational data on the survey administration were collected; for this report, these data are called "paradata" or "process data." The process data collected for the design decisions began at the listing stage of the sample and ended at the last call on the last case in March 2003. The data included the following:

Observations on Each Interviewer

a. Experience in interviewing tasks
b. Performance on training

Lister Observations on the Sample Segment

a. Evidence of abandoned or unoccupied structures
b. Extent of commercial, church, school, and other nonresidential use in the neighborhood
c. Physical access impediments (e.g., walled subdivisions)
d. Evidence of non-English speakers in the neighborhood
e. Evidence of safety concerns for the interviewer

Lister Observations on Each Listed Housing Unit

a. Access impediments to the unit (e.g., locked gates)
b. Number of housing units in the structure
c. Publicly visible evidence of children (e.g., toys visible)
d. Publicly visible evidence of adult at home during the day

Observations on Each Call to the Unit

a. Time of day
b. Day of the week
c. Outcome of call

Observations on Each Call Yielding a Contact with a Household Member

a. Whether the householder asked a question (e.g., "How did you choose my house?")
b. Whether the householder noted it was a bad time to talk (e.g., "We were just sitting down for lunch. Could we talk later?")
c. Whether the householder made a negative statement about the survey request (e.g., "How do I know you're really who you say you are?")

Observations for Each Interviewer Day

a. Number of hours spent traveling to the segment

b. Number of hours spent on administrative activities
c. Number of hours spent attempting screening interviews
d. Number of hours spent attempting main interviews

Each of these data items was designed based on prior studies indicating its relationship to difficulty of contacting sample households or difficulty of persuading them to cooperate with an interview.

What results from such a process data design is the following nested structure of process data, as shown in figure 8:

● Interviewers are the highest level of aggregation
● The sample segment paradata are the second level
● The housing unit is the third level
● Records on each contact attempt (or "call" or "visit") form the lowest level of the paradata structure

Paradata exist for the screening stage and the main interview stage of the interviewing. Needless to say, the complexity of the process data rivals the complexity of the interview data themselves. However, as figure 8 implies, statistical analysis of the data can answer questions like, "What calling pattern on screener interviews predicts the need to make more

callbacks to obtain the main interview?" or, "What interviewers achieve more efficient interview production with different kinds of sample persons?"

The Cycle 6 data collection had three phases. Each of the first two phases used the paradata described previously to identify changes that were necessary in the next phase.

Phase 1 (March 2002–June 2002)—The first phase was based on the largest self-representing primary areas, and one-quarter of the nonself-representing areas. Interviewers for this quarter sample were trained and initial replicate samples released in March 2002. Information was needed on: how well the interviewers could perform the tasks they were assigned; how many hours of effort would be required to obtain interviews; how many hours interviewers would work; and what the eligibility and response rates would be. When the data from this first phase were obtained, it would be possible to determine how many trained interviewers would be needed to complete the study and how many interviews they could be expected to produce within the budget. Process data were collected on this quarter sample and forecasts of effort required for the task were made daily. At the beginning of May 2002, final forecasts were set and most design parameters were fixed.

Phase 2 (June 2002–January 2003)—In June 2002, a second set of interviewers were hired and began work in the rest of the 121 primary sampling areas in the full design. Phase 2 ended on January 29, 2003. This phase formed the majority of the data collection activities. Phase 2 ended with all recruitment efforts completed according to the rules specified in June 2002.

Phase 3 (February 2003–March 2003)—Phase 3 collected data on a subsample of the cases remaining nonrespondent at the end of January 2003. Phase 3 altered the recruitment protocol in a variety of ways to be appealing to the subsampled cases. Phase 3 ended the first week of March 2003.

Phase 1: Using Paradata to Specify Callback Rules (March 2002–June 2002)

A common outcome in household surveys is that the early days of the data collection are quite productive of contacts and interviews, but that the last days of the data collection period are quite inefficient. The current theories about survey participation (36) posit that different sets of influences act on sample persons to determine their likelihood of participation. For some, the topic of the survey is of great interest; for others, the use of an incentive is important; for others, the sponsor or data collection organization evokes interest. As Groves and Couper (12) show, the number of questions and comments by both respondents and interviewers decline over the course of repeated contacts with a sample unit. It appears that as the number of calls and contacts increase over the course of a data collection period, the amount of change in nonresponse bias itself declines. This must be true in part because of the declining percentage of interviews obtained with each additional call. However, this declining change in nonresponse bias also occurs because the amount of change in the causes of the participation decision declines over the course of the study. Most of the reasons for refusing and accepting, and most of the situational factors have been experienced by interviewers and respondents.

Phase 1 used a one-quarter sample of primary areas, a reduced interviewer corps, and unlimited callback rules. During Phase 1, estimates of several key NSFG statistics were computed routinely, such as the proportion of respondents who had never been married, the proportion who had never had a birth, and so on. The staff examined the impact of interviewer effort (as indicated by number of contact attempts) on these key statistics. For example, has two y-axes and two associated plots; one, is the cumulative estimate of the statistic, using all interviews collected on or before that call number. This cumulative graph uses the right y-axis and is very

unchanging in its height. The second plot—corresponding to the left y axis—is a much more variable plot. It is the value of the statistic based on the interviews taken only on a particular call number. As that plot moves to the right, the statistic is based on fewer and fewer cases; for that reason, the estimates (for the 30th, 31st, and 32nd call) become very erratic.

During the course of the data collection period, these statistics were examined multiple times to see when the estimates began to show some stability. The call-specific statistic plot was examined to look for the direction of change in the early calls (i.e., 1–10). When there appeared to be a systematic pattern in the movement of the call-specific estimates, then closer attention was paid to the movement in the cumulative plot to see whether the changes were important substantively. Simultaneously, multivariate models estimated on call-level records and time reports from interviewers tracked the average costs of a call on a sample case.

The conclusion after examining these plots of key statistics over the course of the data collection period was that 10–14 visits to a sample household produced stable cumulative estimates on the vast majority of the key statistics. ("Stability" here was defined as values that would yield the same substantive conclusion.) This analysis during Phase 1 led to the choice of the design option for the later phases that a maximum of 10–14 calls would be made on sample cases. Based on the Phase 1 experience, it is estimated that approximately up to 9 percent of the screener call attempts could be eliminated in Phase 2 and 3 screening. Separate paradata models suggested that marginal time required for each screener call was 4.2 minutes. At the volume of interviewer activities forecasted for this survey, this represented a saving of approximately 800–1,000 interviewer hours for the entire survey.

The key indicators and examples of plots used to monitor their values were:

a. mean number of non-coresidential children among males,

b. proportion of males who had never been married,
c. proportion of males who ever had sex,
d. proportion of males who ever fathered a biological child,
e. proportion of never-married females,
f. mean number of live births among females,
g. proportion of females who have ever been pregnant, and
h. proportion of females who ever had sex.

We present two of the plots as illustrations (figures 9a and 9b).

Phase 2: Tracking of Response Propensities During Phase 2 (June 2002–January 2003)

A "double sample" was used to manage the end of the data collection in a way that would limit costs while reducing nonresponse error.

Two models estimating the propensity of a case to be interviewed on the next call were constructed. These models were discrete hazard models using a variety of paradata. In essence, each call was a separate data record; the data record contained all the lister observations on the segment, all the interviewer observations on the sample unit, and recorded behaviors of the household members that occurred in prior calls. One model predicted the likelihood that a case not yet providing a screener interview would do so on the next call. It had the predictors displayed in table N, with their associated coefficients. The model was built in two steps; first a stepwise procedure using the variables at the segment, unit, and call level was employed. Then the model was respecified using past literature and theories applicable to response propensity. This second step removed some nonsensical features of the stepwise specification.

Table N presents coefficients from a discrete hazard model. The dependent variable is whether (or not) the next visit (or "call attempt") results in a completed screener. The largest predictor of a successful screener is prior contact with the household. The
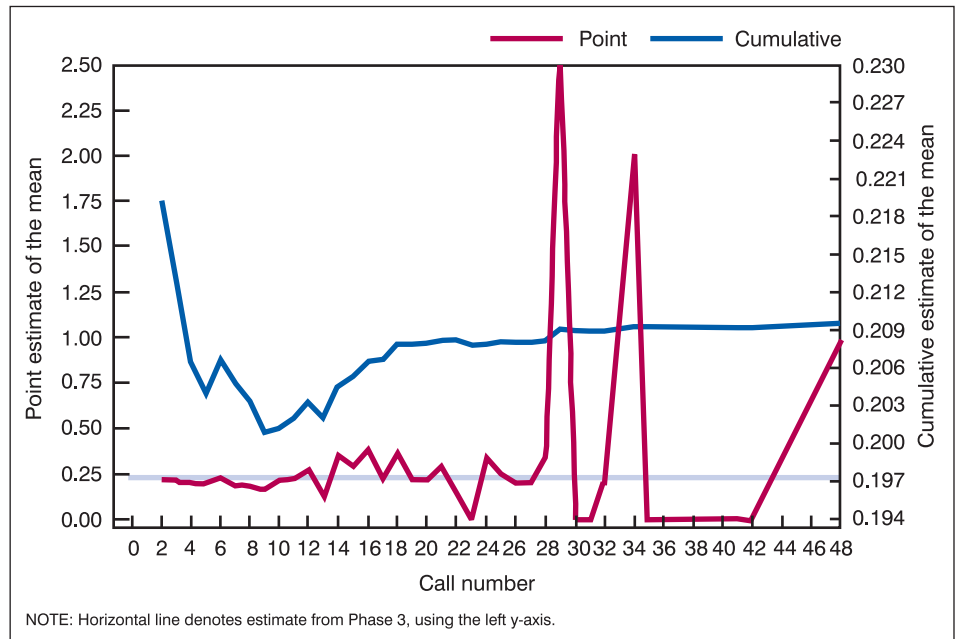
Figure 9a. Mean number of non-coresidential children among males
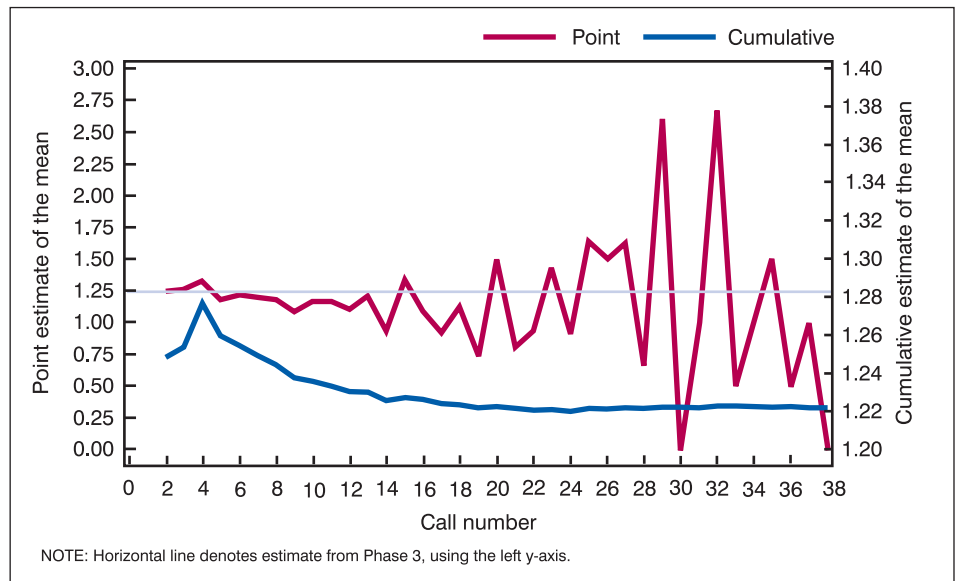
Figure 9b. Mean number of live births among females

Figure 9. Estimates for two key indicators by number of calls required to complete the main interview

variable is placed in the model because of its value to the interviewer in guiding future callbacks. There is also a positive effect if the householder asked some questions about the survey on the last contact (marginal odds ratio=1.06). The latter indicator was derived from results of prior studies that showed that when householders asked questions, they exhibited the kind of engagement that led later to their willingness to respond.

There are relatively large negative effects of the number of prior calls (odds ratio=0.89 for each additional call) and some negative comments in a prior contact (odds ratio=0.24). Access impediments reduce the propensity of an interview in general (odds ratio=0.66). Cases located in large urban areas display lower propensities to respond, a typical finding in household surveys (12).

Table N. Discrete hazard model coefficients and marginal odds-ratios predicting likelihood of completing a screening interview on the next visit or call attempt: National Survey of Family Growth Main Study, 2002

Predictor	Coefficient	p-value	Odds ratio
Intercept	−9.370	<.05	0.00
Urban=1	−0.210	<.01	0.81
Access problems in segment=1	−0.015	NS	0.99
Residential neighborhood=1	0.110	<.01	1.12
Evidence of non-English speakers=1	0.031	NS	1.03
Evidence of safety concerns=1	0.088	<.01	1.09
Evidence of unit-level access impediments=1	−0.420	<.01	0.66
Large multiunit structure=1	−0.100	<.01	0.90
Number of prior calls	−0.120	<.01	0.89
Some prior contact with unit=1	8.740	<.01	6247.90
Number of prior contacts	−0.065	<.01	0.94
Some negative statements by householder=1	−1.420	<.01	0.24
Last contact statements from householder=1	0.650	NS	1.92
Some questions asked in earlier contact=1	−0.085	<.01	0.43
Questions asked in last contact=1	0.056	NS	1.06

0.00 Quantity more than 0 but less than 0.05.
NS Not significant.

Table O. Discrete hazard model coefficients and marginal odds-ratios predicting likelihood of a main interview on the next call attempt: National Survey of Family Growth Main Study, 2002

Predictor	Coefficient	p-value	Odds ratio
Intercept	−3.95	<.01	0.019
Urban=1	−0.12	<.01	0.89
Uninhabited structures in segment=1	−0.037	NS	0.96
Public housing project=1	0.071	NS	1.07
Residential neighborhood=1	0.044	NS	1.04
Evidence of non-English speakers=1	0.023	NS	1.02
Evidence of Spanish speakers=1	0.039	NS	1.04
Evidence of safety concerns=1	0.0081	NS	1.01
Some prior contact with unit=1	4.58	<.01	97.51
Resistance displayed on earlier contact=1	−2.28	<.01	0.1
Large multi-unit structure=1	0.13	<.01	1.14
Evidence of unit-level access impediments=1	−0.088	<.05	0.92
Evidence of security measures in unit=1	−0.016	NS	0.98
Sample person is teenager=1	0.25	<.01	1.28
Sample person is male=1	−0.11	<.01	0.9
Sample person is black=1	−0.017	NS	0.98
Sample person speaks Spanish=1	−0.30	<.01	0.74
Household has only one member=1	0.30	<.01	1.35
Previous call was a contact=1	1.50	<.01	4.48
Some statements by householder=1	−1.43	<.01	0.24
Last contact statements from householder=1	0.71	NS	2.03
Some questions asked in earlier contact=1	0.060	NS	1.06
Questions asked in last contact=1	0.24	<.01	1.27
Number of prior calls	−0.066	<.01	0.94
Number of prior contacts	0.12	<.01	1.13

NS Not significant.

Table O shows a similar display of coefficients for a model predicting the propensity of a main interview on the next call. This model showed that there were strong positive effects of having prior contact with the unit (odds ratio=97.5), the prior call being a contact (odds ratio=4.48), and the number of prior contacts (odds ratio=1.13 for each additional contact). Negative effects of notable magnitude were associated with cases where some resistance was displayed in earlier contact (odds ratio=0.10), whether the householder made some statement during an earlier contact (odds ratio=0.24) (these tend to be negative statements), and the number of prior calls on the case (odds ratio=0.94 for each additional call). This main interview model could examine the effects of person-level characteristics of the selected respondent on the propensity and found the expected positive effects of the respondent being a teenager (odds ratio=1.28), negative effects of being male (odds ratio=0.9), and negative effects of a Spanish-speaking respondent (odds ratio=0.74).

Expected values for each active case, given the model, were computed several times during the data collection period. The expected values were summed over all cases within a sample segment (weighting the screener model expected values by the expected eligibility of the households). Two uses were made of the segment totals of expected values:

1. Segments were grouped into categories with low, medium, and high total propensities, for use by supervisors to direct the work of interviewers to the most promising areas
2. At the end of Phase 2, segments were grouped into quartiles that formed strata for the Phase 3 sample

Phase 3: The Subsample of Nonrespondents (February–March 2003)

As the NSFG Main Study data collection matured, the preparations for the Phase 3 subsample design and recruitment protocol began. "Two-phase sample designs" were first defined by Hansen and Hurwitz (37) as tools to reduce the amount of nonresponse bias in survey statistics. The ingredients of a two-phase design for nonresponse bias reduction are:

1. The design and implementation of a survey design on a given sample (labeled the "first phase")
2. The selection of a probability sample of the nonrespondents to the first phase implementation
3. The use of a different participation protocol for the second phase

There are three impacts of a two-phase design. First, if the second phase protocol is successful in measuring 100 percent of the *sampled* nonrespondents from the first phase, nonresponse bias in all statistics is

Table P. Number of Phase 1 and Phase 2 segments by number of active incomplete cases in the segment at the end of Phase 2, and the estimated total propensity to complete an interview on the next visit: National Survey of Family Growth Cycle 6

Total propensity for active cases remaining	Number of active cases remaining uncompleted at end of Phases 1 and 2 (quartiles)				
	Total	Highest	Medium-high	Medium-low	Lowest
Total.	783	206	201	180	196
High.	195	111	51	26	7
Medium-high	196	55	65	48	28
Medium-low	196	30	51	66	49
Low	196	10	34	40	112

Table Q. Relative sampling fractions in Phase 3 by number of active cases remaining, and the estimated total propensity to respond on the next call: National Survey of Family Growth Cycle 6

Total propensity for active cases remaining	Number of active cases remaining uncompleted at end of Phases 1 and 2 (quartiles)			
	Highest	Medium-high	Medium-low	Lowest
High	0.58	0.58	0.58	0.58
Medium-high	0.58	0.58	0.58/2	0.58/2
Medium-low	0.58/2	0.58/2	0.58/3	0.58/3
Low	0.58/2	0.58/3	0.58/3	0.58/4

eliminated. Second, the cases sampled into the second phase who 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. In Cycle 6, a subsample of nonrespondents (what Hansen and Hurwitz called a second phase sample) was introduced as Phase 3 of the design.

In practice, no subsample of nonrespondents attains a 100 percent response rate and thus some nonresponse bias remains. Given the theoretical perspective guiding the NSFG (36), an attempt was made to mount a recruitment protocol that was distinctive from the Phase 1 protocol. Under the theory, such a protocol would attract effectively those sample persons who had rejected the protocol of the Phase 1 design. The combined set of respondents would thus be more diverse, and thus more representative of the U.S. population, than the respondents from the early phases.

Multiphase designs are increasingly attractive to survey researchers in the United States because they offer a way to control the costs at the end of a data collection period, with concerns about nonresponse errors as well as

nonresponse rates. In most face-to-face surveys, at the end of the data collection period, large costs are spent to travel out to sample segments to visit only one or two sample households, usually those who have proven to be extremely difficult to contact in prior visits or who have displayed reluctance to grant the survey request. By restricting these expensive visits to a sample of the remaining nonrespondents, costs can be saved.

Based on the propensity models above, the 783 sample segments of Phase 1 and 2 were stratified on two major dimensions: the number of cases in the segment that were not finalized and the total expected propensities for active cases in the segment based on the models shown in tables N and O. The term "high" refers to the top quartile, "medium-high," the second quartile; "medium-low," the third quartile, and "low," the fourth or lowest quartile. This resulted in the stratification in table P.

It was believed that the most attractive cell in table P was the cell for segments with high propensity to complete an interview on the next visit, and a high number of active cases, containing 111 segments. When visiting these segments, the interviewers could visit several houses on the same trip.

The Phase 3 sample was a stratified sample of segments, with all nonrespondent cases in a selected segment included in the Phase 3 sample. This was chosen based on cost model estimates computed during Phases 1 and 2 that showed that a large portion of the total cost of completing a case arose from travel cost to the sample segments. Sampling fractions across the strata varied by a factor of 4.0, as shown in table Q. Note that this design option placed large emphasis on the cost efficiency of the Phase 3 design to produce interviews, not on minimizing the standard errors of the resulting data set. However, the variation in sampling fractions over the 16 strata represents a 1 to 4 ratio; simulations prior to the selection suggested an increase in variance due to the additional selection weighting of approximately 20 percent. The highest selection probabilities were assigned to those segments with large total expected propensities to be interviewed or large numbers of active cases. The smallest selection probabilities for the Phase 3 sample were assigned to segments with few active cases that had low propensities of being interviewed, given the previously shown models.

Under the theoretical perspective guiding the NSFG fieldwork, we sought to design a Phase 3 recruitment protocol that was distinctive from that used in Phases 1 and 2. Such distinction is necessary (but not *a priori* sufficient) to attract sample persons who did not find the ingredients of the Phase 1 and 2 protocol effective for them. With the support of two Institutional Review Boards and the Office of Management and Budget, the Phase 3 recruitment protocol involved the following ingredients:

1. Use of the most productive interviewers on staff
2. Increased use of proxy informants for the screening interview
3. A prepaid $5 incentive (versus no incentive) for cases that had not yet completed the screening interview
4. A prepaid $40 incentive for the main interview (compared with no prepaid incentive in Phases 1 and 2)
5. A promised additional $40 incentive for a completed main interview

We limited Phase 3 to a 1-month period, following 11 months in the first two phases.

Ideally, the amount of effort would be more uniform across segments of different numbers of active cases. Such uniformity might produce more equal response rates across the strata.

Table R presents overall average segment screener response rates, defined as the (number of screener interviews) divided by (the number of eligibles) during the 1 month of Phase 3 fieldwork within a segment, averaged over all Phase 3 segments. The overall mean screener response rate per segment was 51 percent, a rather remarkable rate given the fact that the sample had been worked for 12 months in Phases 1 and 2. The response rates range from 0 to 89 percent for the active cases for which screeners were sampled. Ideally, the response rates would be constant over all 16 cells of the Phase 3 design. There is much greater stability of these response rates at the margins, reflecting the greater number of segments on which the means are based. Across the four grouped strata for active cases (the "Total" row of table R), the response rates range from 49 to 55 percent with no trend by number of active cases. For the grouped propensity strata the range is from 49 to 53 percent (the "Total" column of table R), again with no trend by magnitude of propensities. These are desirable properties, given the value of balancing representation of the Phase 3 sample over the 16 strata.

Table S presents the mean per segment main interview response rates during Phase 3 for eligible sample persons within Phase 3. Across the 16 cells the variation in response rates is from 0 to 64 percent. The marginal variation, again, is much smaller. For strata defined by active case count (the "Total" row of table S), the range is from 42 to 54 percent without a trend. For strata defined by total propensities (the "Total" column of table S), the range is from 35 to 54 percent with a trend toward higher response rates in the segments with higher expected propensities. This is undesirable but perhaps a measure of the difficulty of obtaining the main interviews after a prior refusal.

Table R. Mean screener response rates for segments during Phase 3 by number of active cases remaining in the segment, according to total propensity to respond on the next call: National Survey of Family Growth Main Study, 2002

Total propensity for active cases remaining	Number of active cases remaining uncompleted at end of Phases 1 and 2 (quartiles)				
	Total	Highest	Medium-high	Medium-low	Lowest
Total	0.51	0.51	0.49	0.52	0.55
High	0.50	0.52	0.42	0.59	0.00
Medium-high	0.53	0.51	0.51	0.53	0.83
Medium-low	0.49	0.56	0.48	0.41	0.54
Low	0.52	0.28	0.89	0.50	0.33

0.0 Quantity more than 0 but less than 0.05.

NOTE: Average over segments of the ratio of completed screeners in third phase to number of cases not yet screened at beginning of the third phase.

Table S. Mean response rates to the main interview for segments during Phase 3 by number of active cases in the segment, according to total propensity of cases in the segment to respond to the next call: National Survey of Famiy Growth Main Study, 2002

Total propensity for active cases remaining	Number of active cases remaining uncompleted at end of Phases 1 and 2				
	Total	High	Medium-high	Medium-low	Low
Total	0.51	0.54	0.49	0.53	0.42
High	0.54	0.54	0.50	0.64	0.50
Medium-high	0.53	0.56	0.57	0.51	0.31
Medium-low	0.43	0.51	0.35	0.35	0.57
Low	0.35	0.33	0.00	0.65	0.33

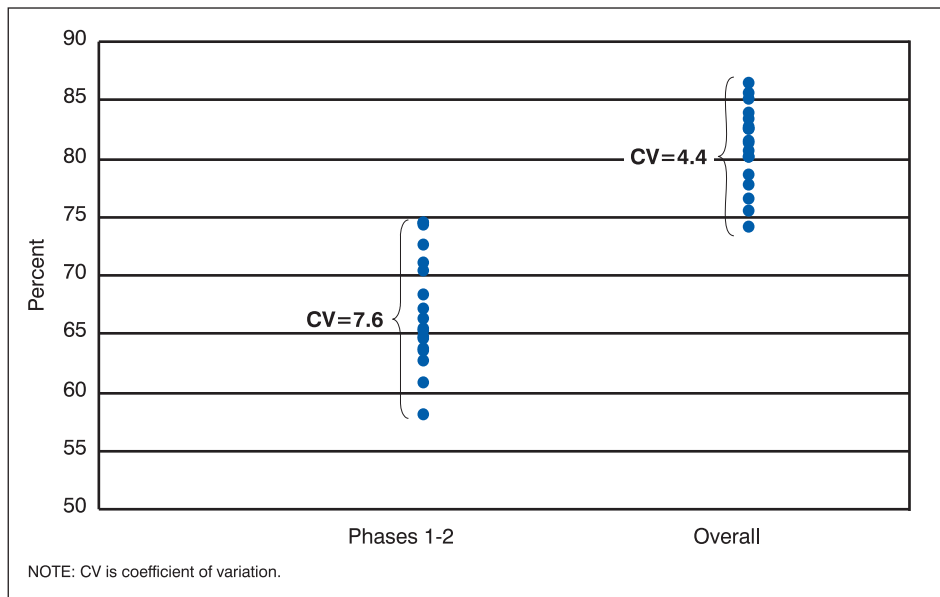0.0 Quantity more than 0 but less than 0.05.

NOTE: Average over segments of the ratio of main interviews in third phase to number of cases eligible for the main interview in the third phase.

Another issue is assessing whether the stratification used in Phase 3 helped to control the costs of data collection in that phase. There is evidence that interviewers made more calls per screener case for segments that had more active cases (on average about four calls in the large segments and three calls in the small segments). There is no consistent pattern of effort over segments with different estimated propensities to respond to the screener. There are smaller differences across sample segments in the effort expended to obtain main interviews.
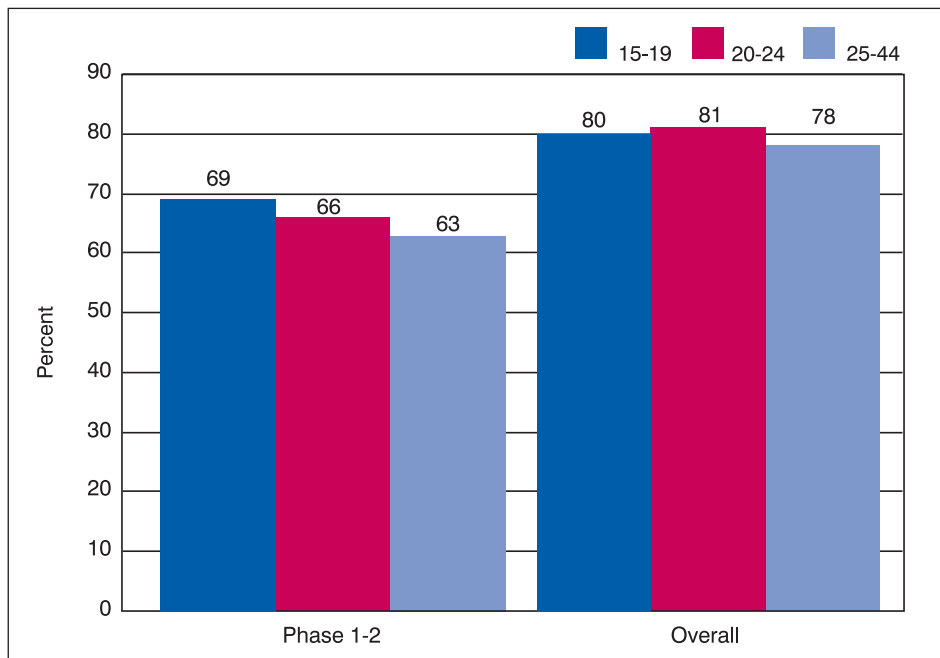
A measure of the cost efficiency of effort is the ratio of interviews to the number of calls required to achieve them; the higher the ratio, the more efficient is the sample to complete. For screener interviewing, the ratio is unexpectedly higher for the low propensity stratum (.24 interviews per call) than the high propensity stratum (.16 interviews per call). For the main interview efforts, the high propensity stratum achieves higher efficiency (.16 interviews per call) than the low propensity stratum (.14 interviews per call). Larger efficiency differences arise due to the segment size stratification. Both for screening and main interviewing, the stratum of small segments achieved higher efficiencies than the stratum of large segments (.34 to .14 screener interviews per call; .20 to .16 main interviews per call). In short, the higher response rates of the high propensity strata do not come at dramatically higher levels of efficiency in terms of callbacks.

The overall response rate at the end of Phases 1 and 2 was approximately 64 percent, using the AAPOR definition that uses an estimated eligibility rate among the nonrespondent screener cases but does not reflect unequal probabilities of selection (38). The Phase 3 response rate was approximately 40 percent, which yielded a combined response rate of between 78–79 percent, using the approved AAPOR double sample computation (again not reflecting unequal probabilities within phases). In

Figure 10. Response rates for Phases 1 and 2, and Phases 1–3 for 18 subpopulations by age, sex, and race: National Survey of Family Growth Main Study, 2002

Figure 11. Response rates for Phases 1 and 2 and for Phases 1–3 (overall, reflecting subsampling of nonrespondents) by age group: National Survey of Family Growth Main Study, 2002

overall response rates of the 18 subpopulations is one way to examine the imbalance of the Phases 1 and 2 performance across these 18 subpopulations, among successfully screened households.

Figure 10 presents the response rates at the end of Phases 1 and 2 and the overall response rates for the 18 subpopulations. The coefficient of variation of the response rates (standard deviation of response rates over the 18 subpopulations divided by the overall response rate) in Phases 1 and 2 is 7.6 percent of the mean response rate; the same measure of variation for the overall rates is 4.4 percent, a large decrease in the variation in response rates. We take this as an indirect indication of reduced nonresponse error associated with age, gender, and race/ethnicity variation.

Figure 11 shows that the variation is quite systematic by age of the sample persons. At the end of Phases 1 and 2, teenagers had a 6 percentage point higher response rate than the oldest age group, with those 20–24 years of age in between the two. At the end of Phase 3, the difference in response rates among the three groups was just 3 percentage points.

Are there indications of the impact on estimates of this reduced variation? One would expect that the impact on estimates of the overall rates would be to impact all the fertility experience variables that are a function of age. Figures like figure 12 were used to present the same estimates as those appearing in figure 9, now with the values of the Phase 3 estimates appearing, juxtaposed to the change in the estimate as a function of call number. On the plot of the call-specific estimate appears a horizontal line that is the value of the Phase 3 estimate (using the left y-axis). On the plot of the cumulative value of the estimate, on the right y-axis appears a single point that is the value of the Phase 3 estimate. As would be expected, the proportion of females who ever had sex among Phase 3 respondents is higher than that of the first phase respondents, because the response rates for adults in Phase 3 were higher than the response rates in Phases 1 and 2, relative to those of younger persons.

that sense, the Phase 3 sample produced a very successful increase in the overall response rate in just 1 month.

What evidence do we have regarding the nonresponse bias characteristics of the multiphase design? As always with nonresponse error, the evidence is only indirect, because we do not know the characteristics of the people who did not respond. The NSFG attempts to control achieved interview counts on 18 different subpopulations defined by three age groups, two gender groups, and three race/ethnicity groups. The screening information provided data on age, gender, and race/ethnicity for successfully screened households. Examining the Phases 1 and 2 and
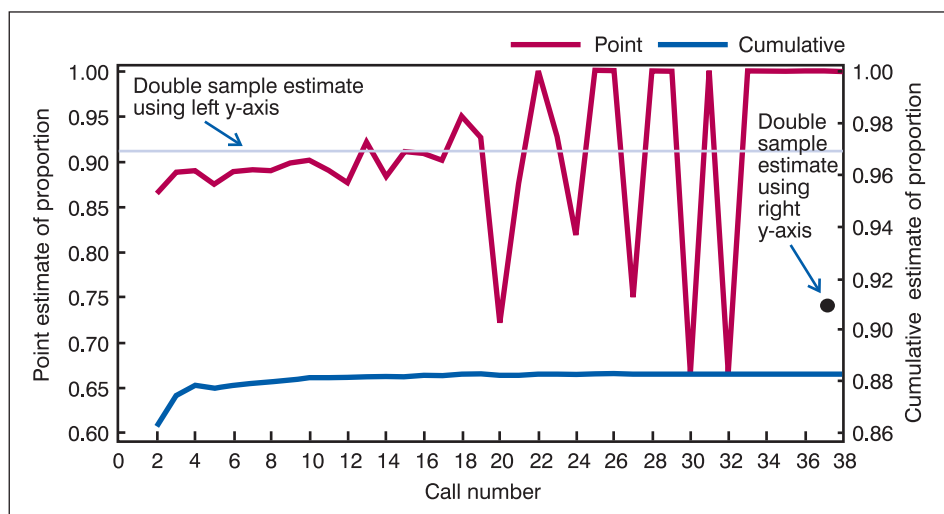
Figure 12. Estimated proportion of females ever had sex by call number of interview, displaying Phase 3 estimates

decline and reach a plateau of about 10.9 hours per interview in the fall months and then begin increasing again, as the interviewers complete the easy-to-contact and easy-to-interview cases and focus on the more difficult. The February 1, 2003, Phase 3 introduction prevented the rapid increase in hours per interview seen in Phase 1 by implementing more efficient protocols, on a subsample of the remaining nonrespondent cases.

Quality Control

Practice, Monitoring, and Evaluation at Training

Quality assurance and control started early in the NSFG Cycle 6 activities. An essential part of the interviewer training involved hands-on practice using the laptop computer and completing the questionnaires. To make sure that the interviewers were properly prepared for their work, ISR trainers and NCHS staff monitored the interviewers during the training session. After the training session each day, the lead and assistant trainers from all the classrooms would meet together with NCHS staff to discuss the overall progress of training and identify any interviewers having trouble following the lessons.

A study hall was held on two or three evenings during each training session for about 2 hours. It was staffed with several trainers and computer support staff. Interviewers who were having difficulty with the material, had a technical problem or question, or simply wanted extra practice were welcome to attend. In some cases, a trainer would encourage a particular struggling interviewer to attend. Many study hall attendees grasped the basic concepts of the interview and the procedures, but wanted additional practice or wanted to go over the materials again at their own pace.

Informally, the trainers observed the trainees continuously, especially during the "roundrobin" interviews. The trainers were able to listen to each interviewer individually and watch how

All of these findings are consistent with the findings of lower response rate variation after Phase 3, by age of the sample persons. Phases 1 and 2 ended with a deficit of respondents older than the teen years. (This may be a function of an enhanced value of the $40 incentive to teenagers in the Phase 1 and 2 design compared with older sample persons.) Phase 3 was more successful in attracting older respondents. The effect of this on key statistics of the survey is that the prevalence of attributes that reflect longer sexually active lives is slightly higher when Phase 3 respondents are included in the sample. From these simple analyses there is evidence of reduced nonresponse error associated with the demographic characteristics because of the Phase 3 design.

Tracking Effort Statistics During the Three Phases of Data Collection

Responsive designs monitor both error-related indicators and cost-related indicators during data collection. A variety of statistical models were used to evaluate the progress of field work and to decide where to allocate interviewer effort in the final stages of interviewing. Those statistical models included:

1. A model linking main interview outcomes with hours spent traveling, doing administrative tasks, making screener visits, and making main interview visits
2. Models measuring the marginal impact on interviewer hours of completed main interviews for the 18 demographic subgroups
3. Models forecasting the final number of interviews under different allocations of staff

These models were used in conjunction with the statistical tracking mentioned previously.

In addition, simple statistics were tracked to identify key lifecycle changes in the data collection. For example, figure 13 shows the cumulative ratio of total interviewer hours worked to total main interviews collected. It has the pattern common to surveys involving screening interviews. The first weeks of Phase 1 especially in April 2002, based on the quarter sample, were devoted mainly to identifying eligible sample persons and relatively few main interviews were obtained. This produced large average hours per interview. Then there was a decline, and a rapid increase toward the end of Phase 1 (the third week of June), when interviewers were mostly making repeated visits on those who were difficult to contact or those reluctant to be interviewed. The beginning of July 2002, saw the release of new sample to staff the full 121 PSU sample with the interviewers, and a repeat of the high hours per interview, as the new cases were contacted and screened. Hours per interview gradually

Figure 13. Cumulative mean number of interviewer hours worked per main interview completed, by day of the data collection period: National Survey of Family Growth Main Study, 2002

well she handled problems and keyed in data. This observation proved informative to share trainers' feedback either during the sessions, during a break, or at the end of the training day.

Certification Interview

A formal evaluation process occurred at the end of training through a face-to-face certification interview. A trainer posed as the respondent and followed a scripted set of answers to each of the questionnaires. The certification interview was designed to duplicate a real interview situation as much as possible. He or she also evaluated the trainee on the following aspects of the interview:

- Explaining the study, selection process, and consent procedures
- Administering the screener
- Addressing respondent questions
- Reading questions verbatim

- Using study specific and general interviewing probes correctly
- Providing feedback to the respondent
- Using the interview aids correctly (Question by Question objectives, Show Cards, Pill Chart, Life History Calendar, and the Family Planning Clinic database)
- Explaining and setting up the ACASI for the respondent

At the end of the certification interview, the training staff person provided oral feedback to the trainee and completed a written, numerical evaluation for ISR. After all of the certifications were completed and while the trainees were returning to their homes, the written evaluations were reviewed by ISR staff. Figure 14 presents the distribution of certification scores. The passing score was 19, as determined by the training staff.

Depending on the certification score, the interviewers were either released to begin work on sample cases or were required to study and complete another certification interview, this time over the telephone. Regardless of her score, if the interviewer had not correctly completed the informed consent process, she was required to redo that part before she was released to work on sample cases. The bilingual interviewers did not complete a face-to-face certification. They were certified over the telephone.

Production Review Interviews

Production review interviews were conducted with each field interviewer several times throughout data collection. These production review interviews were similar to the certification

interview done at the end of training except that they were done over the phone and they used different prepared scripts. Production review interviews were conducted by Team Leaders, always evaluating a member of another team. After a production review interview, the Team Leader would give oral feedback to the interviewer, and send a written copy of her evaluation to the interviewer, the interviewer's Team Leader, and the central ISR office. If the interviewer did not pass the production review interview, she was instructed to stop fieldwork until she had reviewed the problem areas with her Team Leader, practiced her skills, and passed a new production review interview.

Verification

As a quality control measure, a percentage of the cases completed by each interviewer were routinely called back. Occasionally, during these calls, a person did not recall being contacted, or reported some other information that contradicted existing data about the case. Whenever this kind of situation was discovered, it was investigated as a possible falsification, and the Project Manager and Regional Field Managers were contacted immediately. If they confirmed that there appeared to be a problem, they instructed the interviewer to stop working her sample immediately. Additional cases for that interviewer were flagged for verification calls, and these calls were given high priority. If, based on these additional calls and further investigation into the original case, the Project Manager and Regional Field Managers determined there was no falsification, the interviewer was allowed to resume working her sample.

If further evidence of falsification was found, or if there was a high level of nonresponse in verification, all sample lines (i.e., distinct housing units listed prior to the data collection within sample segments) finalized by the interviewer were verified. If any case confirmed that an interviewer had falsified data, the interviewer was immediately terminated. If an interviewer was terminated, all of her completed interviews and age-ineligible households had to be verified. The lines
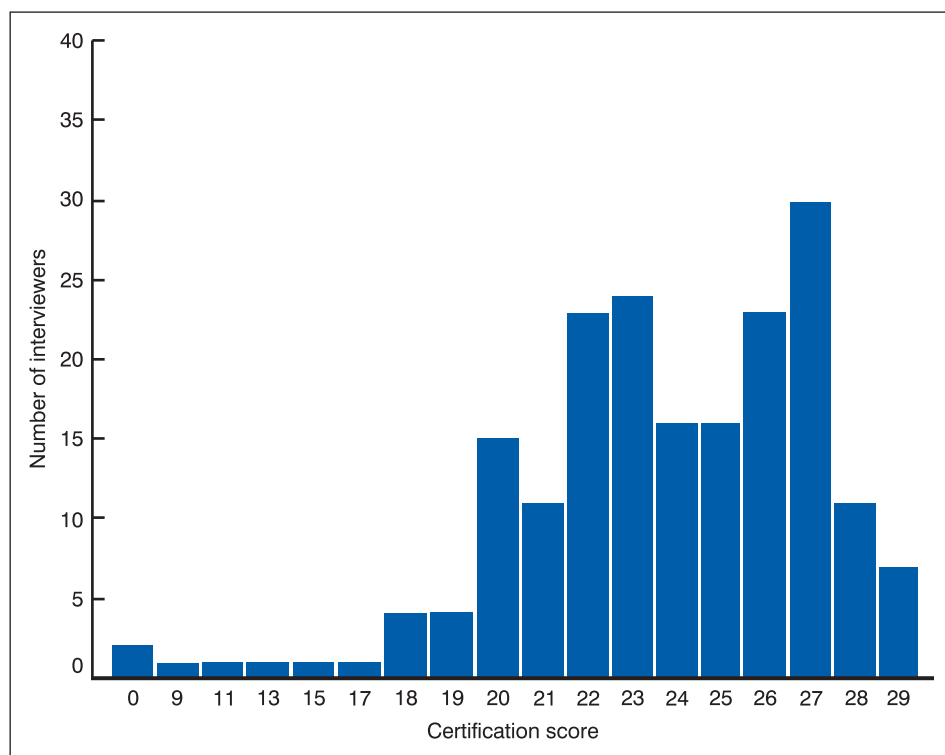
Figure 14. Initial certification scores by interviewer count (passing score, 19): National Survey of Family Growth Main Study Training, 2002

that did not pass verification or had not been verified by telephone were cleared and sent to another interviewer in the field. She verified the lines and redid the ones that failed. Any lines that passed verification were restored with the original data.

The process of verifying fieldwork was a critical component of the overall quality control system. Verifications were completed by telephone for about 1 out of 6 interviews. Following current best practices (39), a probability sample of all interviewers' work was verified. The final verification response rates and discrepancy rates for the subset of cases predesignated for verification are presented in table T. In the table the proportion of interviews verified with some discrepancies reflects those cases where one or more questions on the verification interview had different answers than on the interview, but the central office judged that no falsification had occurred. As part of this process, the staff discovered six interviewers who falsified work. When falsification was discovered, the interviewer's employment was terminated. In all cases, the falsified work was redone by

another interviewer. The rates in table T reflect the status of verification after any falsified cases have been removed and replaced with an actual interview.

Interviewer Comments on Individual Questions

Sometimes during the course of an interview, an issue arose that created a problem or raised a question for the interviewer. Using a preprogrammed key (F2) the interviewer could immediately make a note about the issue. Interviewers recorded such comments for several reasons, including: the interviewer made a self-assessed error and was explaining what she should have done; the software did not permit the interviewer to record the answer she wanted to record (e.g., the response categories seemed to miss the meaning of the respondent's answer); or the interviewer didn't know something that she should have known (e.g., she entered a comment when she could have used an existing code).

The content of these "F2 Comments" from completed interviews was reviewed periodically, summarized and

Table T. Verification response rates among cases predesignated for verification sample, proportion of verified interviews with discrepancies and total verified interviews by instrument: National Survey of Family Growth Main Study, 2002

Instrument	Verification response rate	Proportion interviews verified with some discrepancies	Total interviews verified
Screener total	0.53	0.07	777
Main interview total	0.66	0.03	1,234
Female	0.68	0.03	761
Male.	0.64	0.04	473

evaluated for evidence of application errors or human-computer interface weaknesses. In some cases, this information triggered fixes in the computer application itself and in other cases the information in the comments was used later to make decisions about remedial training or data editing.

Comments made by interviewers were typically clustered within some sections of the questionnaire—for example, Section E for females—as can be seen in the distributions presented in table U. The table shows that the rate of interviewer comments is very low in general, but highest for Section E of the female questionnaire, which contained the detailed contraceptive history (yet even there, only 0.2 percent of the questions in Section E had comments entered). Section E ended with a summary screen, asking interviewers to verify parts of a respondent's sexual, pregnancy, and sterilization experiences. The summary screen review uncovered some problems that originated in earlier questions, which were recorded in Section E.

Section A of both questionnaires contained the household roster, which generated comments on the racial classification question, especially for those respondents who did not want to provide a race for themselves or others in the household.

In general, the comments made by interviewers were of three types. The first type concerned the computer or the interplay between the computer and the survey instrument. Examples of these types of comments include: response options appearing on interview show cards but not on the computer application; questions asking for "all that apply" but allowing only a single entry; question wording being incomplete, in another language, or not making sense; and respondents getting questions they should have been skipped through based on previous answers.

The second type of comment related to clarification of answers when either the respondent or the interviewer wanted to add something to the answer chosen in the survey instrument. Common issues of this type include where the respondent wanted to clarify an earlier report (e.g., date of 1st sex or date of a pregnancy), respondents refusing to choose a racial category, but the computer not allowing the interviewer to proceed without one; male respondents reporting being younger at first sex than the minimum age allowed by the computer; females reporting using contraception (especially pills) for reasons other than birth control; and female respondents changing the order of pregnancies.

The final type of comment was made when the interviewer wanted to report something about the respondent or the circumstances of the interview. In these cases, the interviewer noted issues such as when the phone rang or some other distraction occurred and the interview was suspended for a few minutes.

In addition to providing insight into the problems of individual interviews or interviewers, the list of comments is diagnostic of sections of the questionnaire that may be improved in the next cycle of NSFG, or of issues that may be addressed in the training sessions for the next cycle of the NSFG.

Table U. Number of interviewer comments and percentage of questions with interviewer comments by section of the questionnaire by gender: National Survey of Family Growth Main Study, 2002

Section	Female interviews			Section	Male interviews		
	F2 comments in section	Questions in section	Percent with comments[1]		F2 comments in section	Questions in section	Percent with comments[1]
A	624	43	0.19	A	381	45	0.17
B	593	112	0.07	B	53	45	0.02
C	717	137	0.07	C	145	87	0.03
D	267	61	0.06	D	369	108	0.07
E	1,587	95	0.22	E	110	61	0.04
F	408	43	0.12	F	106	44	0.05
G	51	13	0.05	G	191	38	0.10
H	328	44	0.10	H	38	11	0.07
I	387	56	0.09	I	324	38	0.17
J	46	103	0.01	J	329	52	0.13
Other.	2	NA	NA	K	30	114	0.01
				Other	49	NA	NA

NA Not available.

[1]Number of comments / [(number of questions in section) × (number of interviews)], where number of interviews=7,643 females and 4,928 males.

Interviewer Documentation of Respondent Reactions to Interview

The NCHS Institutional Review Board requested that the NSFG continue a practice used in the pretest: to collect data on whether respondents became "upset" by the main study interview. To do this, and to give project staff additional data on respondent reactions to the interview, the pretest "Interview Observation Form" was revised, and completed by interviewers while respondents were conducting the ACASI component of the interview. Interviewers reported that over 78 percent of the respondents were "friendly" toward the interviewer versus "hostile" or "neither hostile nor friendly." The training of interviewers and the survey materials emphasized respect for the privacy of respondents on the sensitive topics discussed during the interview. These reactions are consistent with the hypothesis that respondent reactions to the interview and the interviewer were generally positive.

One of the observation questions asked whether the respondent became upset during the interview. The interviewer observation form revealed that 1.6 percent of respondents (197 of 12,571) got upset to one degree or another about topics related to the content of the interview, and 0.7 percent (or 93) were upset about subjects unrelated to the interview content. Only 0.4 percent of the respondents (or 48 out of 12,571) remained upset at the end of the interview. None of these events was serious enough to cause the respondents to ask for assistance, or to call project staff in Michigan.

Statistical Process Control Analysis of Interviewer Performance Indicators

As noted above, the Cycle 6 data collection attempted to focus management resources on the issues facing the data collection using quantitative assessments of the production. Key to this focus was the use of statistical process control principles. These principles were applied by viewing interviewers as replicate production units, assigned similar task burdens, and given similar resources to complete those tasks. Variation in the effort required to complete screener and main interviews or variation in the rate of producing successful interviews was studied with ongoing statistical process control charts, like those arrayed in figures 15, 16, and 17. Statistical process control charts were constructed for several indicators on each interviewer: the percentage of the screener sample workload that was finalized; the percentage of the eligible sample cases that was finalized; the average number of hours spent screening for each screener interview produced; and the average number of hours spent attempting main interviews for each main interview produced. All of these statistics were thought to vary by the urbanicity of the sample, access impediments, eligibility rates, and a host of other variables. These sources of variation merely cloud the assessment of interviewer-caused variation. Hence, variation was examined within interviewer teams, which tended to have more homogeneous burdens and shared a supervisor.

Figures 15, 16, and 17 provide a sample of these materials. In figure 15 there are two horizontal lines defining statistical limits to the hours per screener interviewer for a specific team. These limits represent one standard deviation based on among-interviewer variation in average hours per completed screener interview. The lower limit is about .75 hour per screener interviewer; the upper limit is approximately 9.5 hours per screener interview. The last interviewer has the highest hours per successful screener, above that expected given the among-interviewer variation (figure 15a).

Team leaders were instructed to identify the outlier interviewers and then to consider what reasons might explain their high or low productivity.
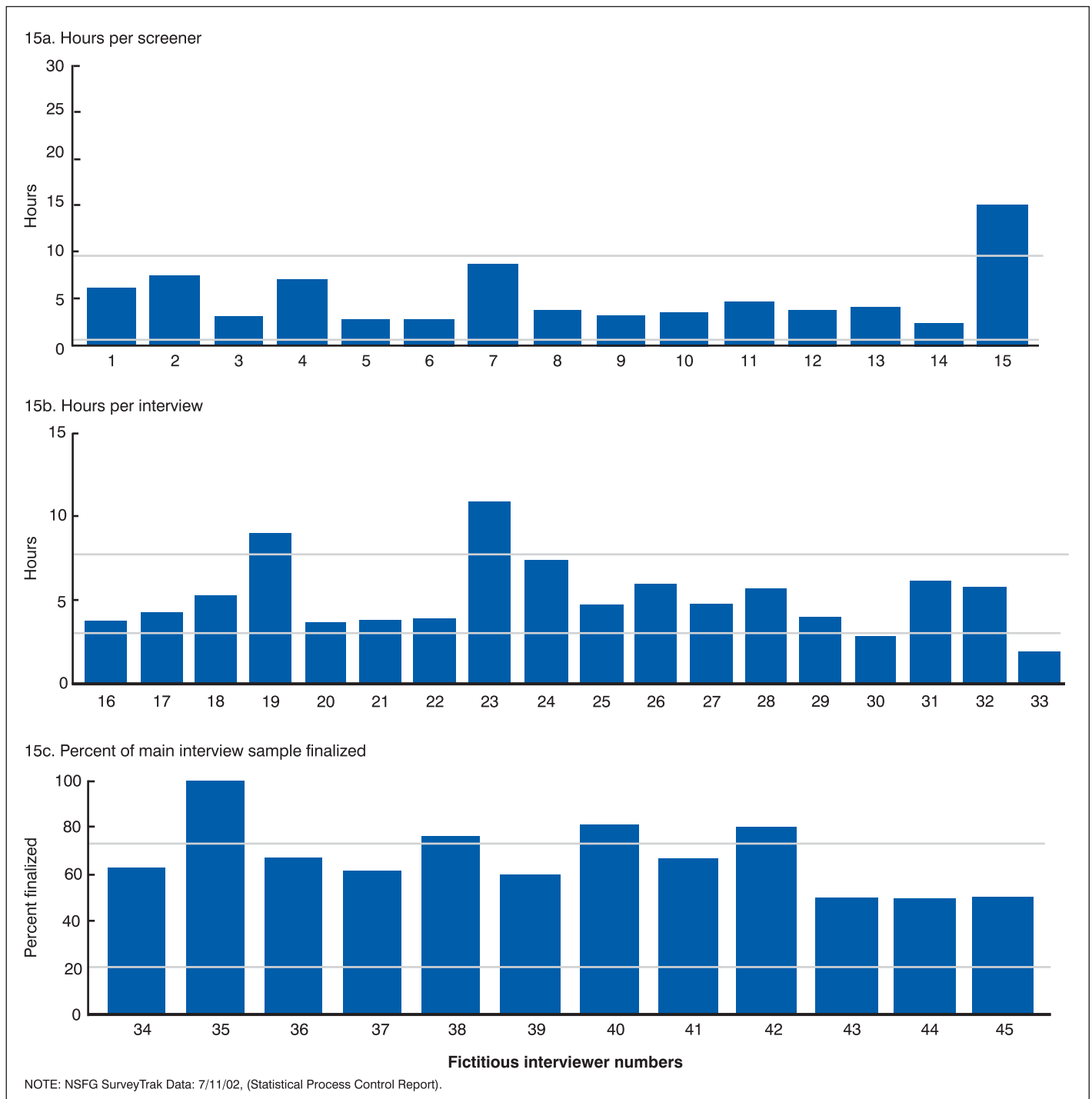
The key tool for examining causes of unusually high or low productivity was the "outlier report," which contained many more statistics describing the work of the interviewer, including whether they were working during hours of the day when persons tend to be home and how their noncontact and refusal rates compared with the average rates. These statistics, with additional expert knowledge of the workload and of the interviewer, guided the supervisor in discussions with the outlier interviewers (see figure 16).

The purpose of the statistical process control input to supervisory interventions was to free supervisors from equal attention to all their interviewers, and to provide an inexpensive data-based rule to greater attention to those interviewers who needed it. Reduction in variation across interviewers was induced both through remedial training but also through dismissal of interviewers who were outliers on the low-end of productivity. What results from this two-pronged effort is a time series of standard deviations in interviewer productivity that changes over time as a function of remedial training and the dismissal of interviewers, often immediately after the June and September interviewer training sessions. Figure 17 displays the standard deviation in hours per main interview for urban area teams and rural area teams. There is reduction in variation in July and October, following dismissals of low performing interviewers and introduction of new interviewers. After October, both groups achieve standard deviations that are rather similar, about 2 hours per interview.

Implication of Lister Observations on Effort Required to Make First Contact with the Sample Household

Given prior research (12), it was expected that listers' observations would be informative about the difficulty interviewers would later have in contacting members of sample households. Thus, interviewers were instructed that, whenever feasible, they should make their first visits to segments where such difficult-to-contact units were found, thus attempting to

15a. Hours per screener

15b. Hours per interview

15c. Percent of main interview sample finalized

Fictitious interviewer numbers

NOTE: NSFG SurveyTrak Data: 7/11/02, (Statistical Process Control Report).

Figure 15. Statistical process control charts

avoid a long period at the end of the data collection period of calls on cases with no prior contact.

Figure 18 shows the measured effects of these attributes on interviewer productivity. The y-axis displays the percentage increase in the number of attempts per first contact. The measure is the total number of calls (attempts) made to cases before making contact divided by the total number of contacts. The average number of attempts per first contact was 2.6 for the total sample. If the sample unit was a single family structure, the average number of attempts to gain one contact was about 10 percent less (or about 2.3 calls). In contrast, housing units in multiunit structures require more calls to first contact, between 20 and 25 percent more than the typical unit.

Similarly, figure 19 shows the impact of various lister observations about the sample unit itself. All the access impediments (a locked central entry point, a locked gate, a doorperson, an intercom at the central entrance) are associated with 25–40 percent increases in the number of calls to obtain a

Interviewer (TLID/IWID)	Peak Call Hrs. %	Non-Peak Call Hrs. %	Screener Non-Contact Rate	Main Non-Contact Rate (overall)	Main Non-Contact Rate (female)	Main Non-Contact Rate (Male)	Screener Refusal Rate	Main Refusal Rate (overall)	Main Refusal Rate (female)	Main Refusal Rate (male)
Average	34.9	65.1	0.19	0.08	0.07	0.10	0.05	0.05	0.05	0.05
201230	46.9	53.1	0.30	0.17	0.10	0.33	0.13	0.10	0.15	0.00
201286	19.3	80.7	0.43	0.28	0.30	0.17	0.07	0.00	0.00	0.00
201314	17.4	82.6	0.00	0.03	0.06	0.00	0.01	0.03	0.00	0.05
201305	46.7	52.3	0.38	0.00	0.00	0.00	0.09	0.00	0.00	0.00
201321	77.6	22.4	0.13	0.40	0.00	0.67	0.09	0.00	0.00	0.00
218599	16.2	83.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
218538	0.00	100.0	0.04	0.00	0.00	0.00	0.01	0.00	0.00	0.00
218592	14.5	85.5	0.18	0.00	0.00	0.00	0.05	0.00	0.00	0.00
218598	49.4	50.6	0.25	0.00	0.00	0.00	0.01	0.00	0.00	0.00
242228	36.8	63.2	0.18	0.04	0.05	0.00	0.04	0.13	0.14	0.13
242204	24.6	75.4	0.34	0.19	0.29	0.00	0.11	0.08	0.08	0.08
242280	30.3	69.7	0.31	0.11	0.13	0.10	0.13	0.06	0.13	0.00
242212	51.5	48.5	0.11	0.07	0.11	0.00	0.12	0.00	0.00	0.00
251384	21.1	78.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
251387	13.0	87.0	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00
261403	32.3	67.7	0.22	0.00	0.00	0.00	0.04	0.00	0.00	0.00
261488	21.3	78.7	0.64	0.07	0.06	0.11	0.05	0.04	0.06	0.00

NOTE: NSFG SurveyTrak Data, 07/11/02.

Figure 16. Example of an outlier report (interviewer numbers are fictitious)

contact. Interviewers, in collaboration with their Team Leaders, gave priority to these cases, in an effort to improve the response rates.

All of these indicators were helpful in two ways: (a) they were used by the field managers to guide the attention of interviewers to those units that were likely to require more effort to contact, and (b) they became useful predictors in propensity models for the likelihood of the interview.

Response Rates

The target response rates for the NSFG were 80 percent for females and 75 percent for males, using the traditional response rates that do not reflect any of the unequal probabilities of selection of the sample design. The Cycle 6 design focused on 18 subclasses that were of key analytic interest, defined by a cross-classification of three age categories, three categories of race/ethnicity, and two categories of gender. Response rates for each of these groups were monitored throughout the data collection period.

There are two complications for computing response rate estimates in the Cycle 6 survey:

1. The sample of housing units required a screening step to determine whether anyone 15–44 years of age was a member of the household. While about 60 percent of the households screened had one or more eligibles, sample units that were not successfully screened had unknown eligibility statuses.
2. A subsample of remaining nonrespondents in Phases 1 and 2 was drawn to represent all nonrespondents.

The American Association for Public Opinion Research guidelines (38) for computation of response rates specifies the following computational form for response rates in such cases. First some notation:

Phase 1 and 2 cases:

s_{i1} = 1, if the case is a screened eligible household; 0, otherwise;

e_{i1} = 1, for eligible households with no screener; 0, otherwise;

u_{i1} = 1, for households of unknown eligibility; 0, otherwise.

Phase 3 cases:

s_{i2} = 1, if the case is a screened eligible household in Phase 2; 0, otherwise;

e_{i2} = 1, for eligible Phase 2 households with no screener; 0, otherwise;

u_{i2} = 1, for eligible Phase 2 households of unknown eligibility; 0, otherwise;

Phase 1 and 2 cases:

i_{i1} = 1, for complete main interviews; 0, otherwise;

m_{i1} = 1, for nonresponse main interview; 0, otherwise.

Phase 3 cases:

i_{i2} = 1, for completed main interviews in Phase 2; 0, otherwise;

m_{i2} = 1, for nonresponse main interviews in Phase 2; 0, otherwise.

The Phase 1 and 2 response rate not reflecting household and person selection weights is a product of a screener response rate (estimating number of screener interviews divided by the number of eligible households sampled) and a main interview response rate (estimating the number of main

interviews divided by the number of eligible sample persons):

$$R_1 = \left(\frac{\sum(s_{i1}+s_{i2})}{\sum[(s_{i1}+s_{i2}) + (e_{i1}+e_{i2}) + l(u_{i1}+u_{i2})]}\right) \bullet$$

$$\left(\frac{\sum i_{i1}}{\sum(i_{i1}+m_{i1})}\right)$$

where l in the denominator of the response rates is an estimate of the proportion of screener nonresponse cases of unknown eligibility that do contain one or more eligibles.
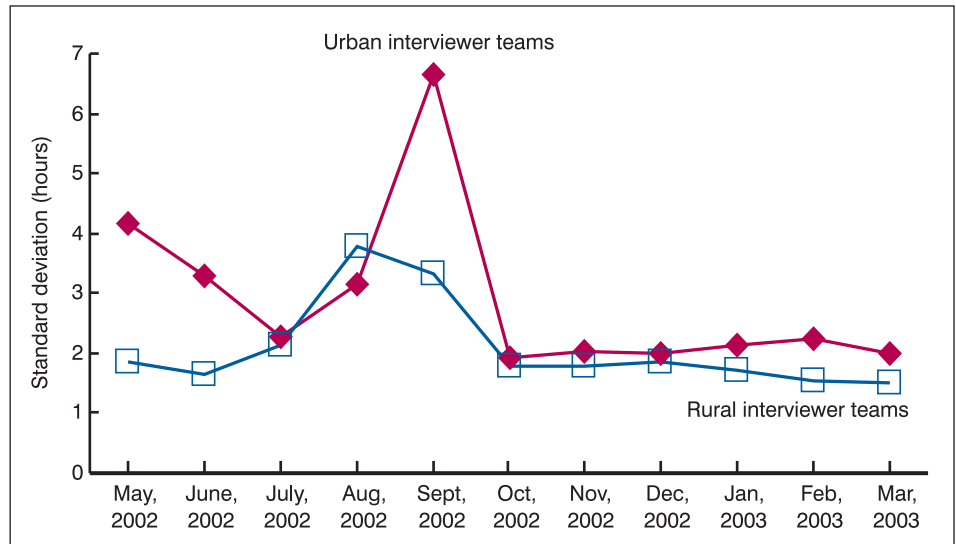
The Phase 3 response rate not reflecting household and person weights is similar, a product of a screener response rate in Phase 3 (estimating number of screener interviews divided by the number of eligible households sampled into Phase 3) and a main interview response rate (estimating the number of main interviews divided by the number of eligible sample persons in Phase 3):

$$R_2 = \left(\frac{\sum s_{i2}}{\sum[s_{i2} + e_{i2} + l(u_{i2})]}\right) \bullet$$

$$\left(\frac{\sum i_{i2}}{\sum(i_{i2} + m_{i2})}\right)$$

We have estimated l above in two different ways:

1. Using the proportion of completed screening interviews that yielded households of one or more eligibles; this is approximately 58 percent
2. Using the estimated proportion with one or more eligibles using a statistical model reflecting the amount of effort required to contact the case and the tendency to be reluctant to the screening interview; this is approximately 64 percent

It is standard practice to present response rates that ignore differential probabilities of selection, an ''unweighted'' response rate. An equivalent estimator of response rates would reflect the subsampling of nonrespondents at the end of Phases 1 and 2 but not the variable selection probabilities. Thus, the response rate for Phases 1 and 2 combined reflects no differential selection probabilities. The

Figure 17. Standard deviation across interviewers in hours per main interview for urban vs. rural teams

Phase 3 response rate is estimated as similarly unweighted, but the two response rates are combined by reflecting the fact that the Phase 3 response rate applies only to the group of cases sampled into the Phase 3 subsample. The overall response rate reflecting the fact that the Phase 3 response rates were based only on a subsample of the remaining nonrespondents of Phases 1 and 2 is:

$$R_1 + (1-R_1)\, R_2$$

The ''unweighted'' response rates computed in this fashion appear in table W.

The response rate estimator above is consistent with prior practice in the NSFG, applied to the multiphase design of Cycle 6. The estimator reflects the sampling of a portion of the segments in the Phase 3 subsample, but does not reflect the unequal probabilities of selection applied to housing units and persons.

The response rate estimators above assume that the screener response rate is the same for all age-by-gender-by-race/ethnicity groups in the design. With that assumption, we can estimate response rates for the 18 key subpopulations in Cycle 6 of the NSFG. Table W shows that the combined response rates for males and females in NSFG Cycle 6 are 78–79 percent, depending on how one estimates the age by gender by race

distribution of households with incomplete screeners.

Looking at the panel of table W that assumes that 58 percent of screener noninterviews contained an eligible person, the overall response rate was 79 percent; the rate for males was 78 percent, and for females, 80 percent. This difference is smaller than expected. For black persons, the overall response rate was 82 percent, for Hispanic persons 81 percent, and for white and other persons, 77 percent. Response rates for teenagers were 80 percent for females and 81 percent for males. The highest response rates were among black and Hispanic teenagers—as high as 88 percent among black female teenagers—and lowest among white (and other) males 25–44 years of age (74 percent).

Data Processing

Data Coding Procedures

A series of programmed computed variables that used respondents' answers to create new variables within the Blaise instrument reduced the need for extensive coding. This was critical in a data file with several thousand variables. Some variables, such as those reporting the contraceptive method history, required staff attention because of the
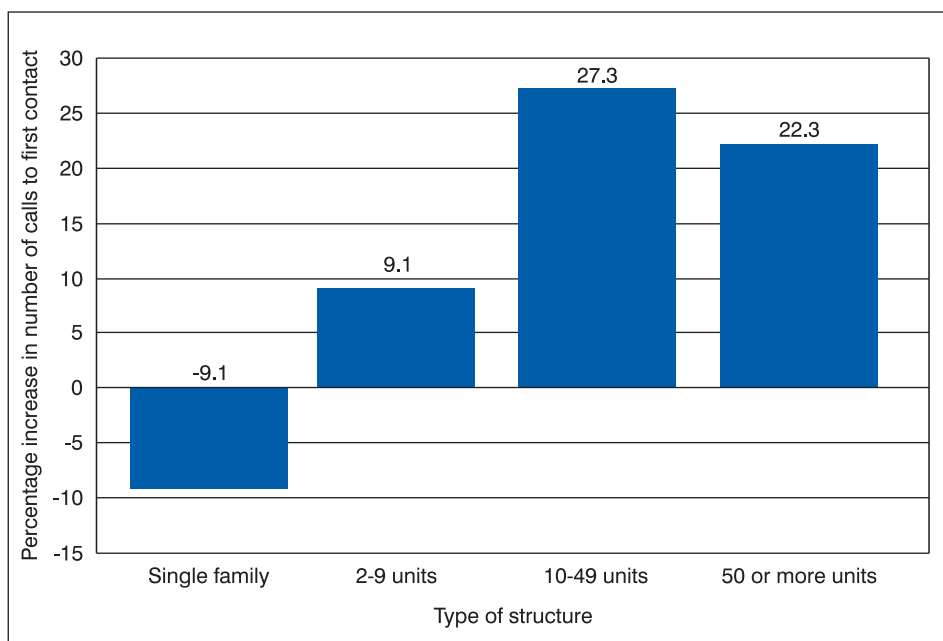
Figure 18. Percentage increase in number of calls before first contact relative to average by type of housing structure
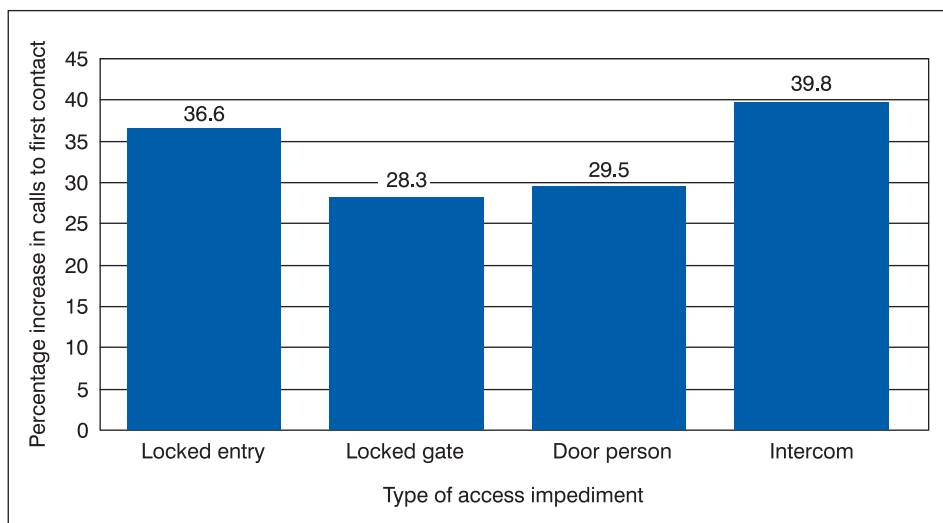
Figure 19. Percentage increase in calls before first contact (relative to total sample) by access impediment type

need to code the "other (specify)" category for responses that did not match the predetermined categories. The answers to questions containing "other (specify)" were recorded verbatim by the interviewer. Such answers were recoded into numeric categories specified by NCHS.

Data File Creation

Upon completion of field interviewing, ISR project staff produced initial versions of the full male and female data files directly from the Blaise data models. These files included all responses captured through the main sections of each questionnaire, the computed variables, and answers provided by each respondent during the ACASI portion of the instrument.

NCHS provided specifications to create 104 recoded variables that would appear in the male file and 399 recoded variables that would appear in the female and pregnancy interval files. These specifications were converted into SAS programming code. The output of

all the recoded variables was extensively examined by NCHS and ISR staff for conformity to specifications and for internal consistency. It was often necessary to produce several drafts of the programs and specifications before the output passed all consistency checks, because many of the recodes were complex in various ways. For example, the structure of the male interview was entirely new, and the female interview was significantly revised from Cycle 5. As in many complex surveys, missing data sometimes led to the discovery of data quality issues that had to be investigated and resolved. The output of all the recoded variables was examined for conformity to specifications and for internal consistency. When necessary, specifications or programs for the recoded variables were revised, until the output passed all consistency checks.

Upon completion of the recode process, each recode variable was subject to an imputation procedure if there were cases with missing data on that recode. A major part of the work of imputation on a file as complex as the NSFG is to (a) specify the conditions under which a variable should be imputed, (b) specify the range of values that can be imputed, and (c) check the imputed values for consistency with other, related variables for the same individual. This specification and evaluation was done collaboratively and iteratively by ISR and NCHS staff. For each variable with imputed cases, an imputation flag was set to indicate that a value was imputed for that case (1).

A special program created a new set of variables in the male respondent file: a presentation of each respondent's biological children in chronological order. The information was not originally collected in this order, but it was deemed necessary to provide it chronologically for data users. Additional variables including weights, time stamps, and selected variables from the screener files were also incorporated into the data file.

The basic layout of the three data files is as follows:

- Respondent ID
- Screener variables

Table W. Cycle 6 unweighted response rates, reflecting subsampling in Phase 3 but no other probabilities of selection by selected characteristics: National Survey of Family Growth Main Study, 2002

| | | Overall Cycle 6 response rates (using .64 eligibility for screener noninterviews) | | | | | | | | |
| | | Female | | | | Male | | | |
Characteristic	Total	All	Black	Hispanic	Other	All	Black	Hispanic	Other
Age									
Total	78	79	81	82	77	77	82	78	75
15–19	79	78	87	84	74	80	83	85	78
20–24	80	80	82	82	79	79	80	80	79
25–44	77	79	80	81	77	75	82	76	73
Race and origin									
Black	82	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Hispanic	80	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Other	76	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
		Overall Cycle 6 response rates (using .58 eligibility for screener noninterviews)								
		Female				Male			
	Total	All	Black	Hispanic	Other	All	Black	Hispanic	Other
Age									
Total	79	80	82	83	78	78	83	79	76
15–19	80	79	88	85	75	81	84	86	79
20–24	81	81	82	83	80	80	81	81	80
25–44	78	80	81	82	78	76	83	76	74
Race and origin									
Black	82	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Hispanic	81	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Other	77	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

. . . Category not applicable.

- Raw and computed variables from sections A through J
- Recodes and Imputation Flags: Sections A-J
- Weights and related variables
- Date of interview

Female (Pregnancy) Interval File

- Respondent ID
- Pregnancy order
- Raw and computed variables for pregnancies from Sections B and E
- Pregnancy-based recodes for sections B and E
- Selected respondent characteristics (raw and recodes)
- Imputation flags for all recodes in this file
- Weights and related variables
- Date of interview

Male respondent file

- Respondent ID
- Screener variables
- Random variable for question-order experiment on contraceptive use
- Raw and computed variables from sections A through K
- Chronologically arranged Biological Child variables
- Recodes and Imputation Flags: Sections A-K
- Weights and related variables
- Date of interview

Data File Documentation

Public-Use File documentation for Cycle 6 is available to researchers as a Web-based tool (called ''WEBDOC'') to permit easy access to all variables, quick navigation between different sections of the instrument, and the capability to search for key concepts and questions. Information on how to obtain the public-use data and how to access the web-based documentation is on the NSFG Web site at http://www.cdc.gov/nchs/about/major/nsfg/nsfgdoc.htm/.

In brief, the public-use data may be downloaded from the NSFG Web site, or obtained free of charge on a CD-ROM by contacting NSFG@cdc.gov.

The web documentation was assembled from various working documents prepared jointly by ISR and NCHS staff: question text, short labels for each variable suitable for use in many statistical software packages, and specifications that defined which respondents were asked each question and which respondents were not. These documents gave an accurate up-to-date view into the data as it went through cleaning, recoding, and imputation. The evolving documentation also served a second purpose—to verify consistency between the study's various products as the dataset was modified.

Key to the construction of this documentation tool was a set of spreadsheets that listed all variables that were output from the Blaise instrument and were used as the driving documents to determine the order and disposition of variables for the public-use file. SAS

data definition statements produced from the Blaise data model were the source of the category labels. A number of scripts were used to load and assemble the documents into a database that underlies the Web documentation.

Two of the most time-consuming tasks in the data documentation phase of the project were the preparation of the ''applicable'' specifications, and the reconciliation of inconsistencies between the various source documents. The first of these tasks, the preparation of ''applicable specs,'' or universe specifications, was done by hand, as in Cycles 1–5. Two different versions of the universe statements from the Blaise instrument and information from internal flow checks were compared and combined by NCHS staff to create a readable version of these specifications.

The second task of reconciliation was greatly aided by inclusion of all the

metadata within a database. This allowed Structured Query Language (SQL) queries to identify common problems such as wild codes, missing labels, negative computed values, variable name misspellings or omissions, etc. These errors would be corrected in the source documents and reloaded into the database. After initially assembling all the basic pieces, additional information segments such as recode specifications and column locations were integrated as they became available.

As part of the documentation package, a User's Guide provides detailed information on the conduct of the survey and on topics such as recoding, imputation, data quality, sample design, estimation procedures, variance estimation, and descriptions of data quality issues for some variables. A variable index is also included. The

Guide is provided as a supplementary part of the documentation but is directly linked and accessible through WEBDOC.

Users access WEBDOC at a level that allows them to view the overall structure of each of the three principal public-use files: male, female, and pregnancy interval. They can examine a list of all the major sections and subsections of each questionnaire and finally, all of the variables included in each subsection.

A sample page (figure 20) of the documentation includes the type of information that is available for each variable. Each recode variable has a link to the specification used to produce it. ''Previous'' and ''next'' buttons allow users to navigate among variables within a section and universe statements include links that allow users to follow the general question flow.
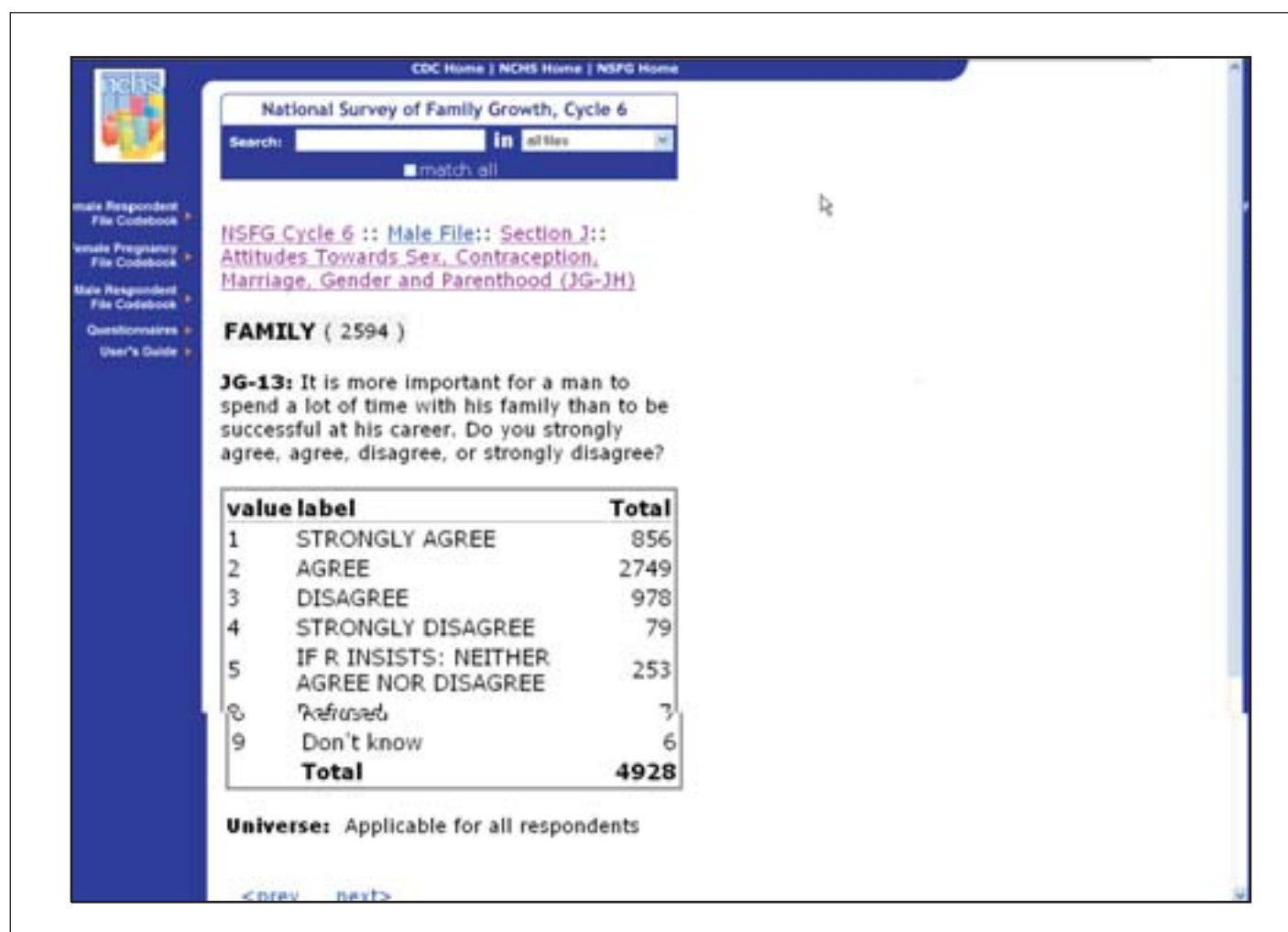

Figure 20. Example of a variable selection from the web documentation

Conclusion

In summary, the effort to ensure the quality of NSFG data began with testing the CAPI programs and continued with interviewer training, verification of interviews, consistency checking (editing), construction and editing of recodes, the evaluation of imputation, and the preparation and editing of the documentation. The intent of all of this effort was twofold: first, to assess and improve the quality of the data for this cycle and for the next cycle of the NSFG; and second, to make the data file and documentation as self-contained, self-explanatory, and easy to use as possible.

References

1. 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). 2005. Forthcoming.

2. Ryder NB, Westoff CF. Reproduction in the United States, 1965. Princeton, NJ: Princeton University Press. 1971.

3. Westoff CF. The yield of the imperfect: The national fertility study. Demography 12(4):573–80. 1975.

4. Mosher WD. Fertility and family planning in the 1970s: The National Survey of Family Growth. Fam Plann Perspect 14(6):314–20. 1982.

5. Mosher WD. Fertility and family planning in the United States: Insights from the National Survey of Family Growth. Fam Plann Perspect 20(5):207–17. 1988.

6. Mosher WD, Bachrach, CA. Understanding U.S. fertility: Continuity and change in the National Survey of Family Growth, 1988–95. Fam Plann Perspect 28(1):4–12. 1996.

7. Mosher WD. Design and operation of the 1995 National Survey of Family Growth. Fam Plann Perspect 30(1):43–6. 1998.

8. Mosher WD, Deang L, Bramlett, MD. Community environment and women's health outcomes: Contextual data. National Center for Health Statistics. Vital Health Stat 23(23). 2003.

9. De Leeuw E, De Heer W. Trends in household survey nonresponse: A longitudinal and international comparison. In: Groves R, Dillman DD, Eltinge J, Little R, eds, Survey nonresponse, 41–54. New York: John Wiley and Sons. 2002.

10. Tourangeau R, Shapiro G, Kearney A, Ernst L. Who lives here? Survey undercoverage and household roster questions. J Official Stat (13):1–18. 1997.

11. Smith T. The hidden 25 percent: An analysis of nonresponse on the 1980 general social survey. Public Opinion Q (47):386–404. 1983.

12. Groves RM, Couper MP. Nonresponse in household surveys. New York: John Wiley and Sons. 1998.

13. Singer E. The use of incentives to reduce nonresponse in household surveys. In: Groves R, Dillman D, Eltinge J, Little, R, eds, Survey nonresponse 163–77. New York: John Wiley and Sons. 2002.

14. Berlin M, Mohadjer L. Waksberg J.et al. An experiment in monetary incentives. In: American Statistical Association, Section on Survey Research Methods, 1992 Proceedings.

15. Ventura SJ, Abma JC, Mosher WD, Henshaw S. Estimated pregnancy rates for the United States, 1990–2000: An Update. National Vital Statistics Reports vol 52 no 23, Hyattsville, MD: National Center for Health Statistics. 2004.

16. Turner C, Forsyth B, O'Reilly J, et al. Automated self-interviewing and the survey measurement of sensitive behaviors. In: Couper M, Baker R, Bethlehem J, et al., eds, Computer assisted survey information collection 455–73. New York: John Wiley and Sons. 1998.

17. Tourangeau R, Smith T. Asking sensitive questions: The impact of data collection, question format, and question context. Public Opinion Q 60:275–304. 1996.

18. Schuman H. The random probe: A technique for evaluating the validity of closed questions. Am Soc Rev 21:218–22. 1966.

19. Tourangeau R, Rips L, Rasinski K. The psychology of survey response. Cambridge: Cambridge University Press. 2000.

20. Rendall M. Clarke L, Peters H, et al. Incomplete reporting of men's fertility in the United States and Britain. Demography 36(1):135–44. 1999.

21. Santelli J, Lindberg L, Abma J, et al. Adolescent sexual behavior: Estimates and trends from four nationally representative surveys. Fam Plann Perspect 32(4):156–65. 2000.

22. Ramirez RR. We the people: Hispanics in the United States. Census 2000 Special reports, No CENSR-18. U.S. Census Bureau, Washington: 2004.

23. Shin H, Bruno R. Language use and English-speaking ability: 2000. Census 2000 Brief, Series C2KBR-29. U.S. Census Bureau, Washington. 2003.

24. Martinez G, Marin B, Schoua-Glusberg A. The back of the tapestry: Translation and administration of the National Survey of Family Growth. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ. 2004.

25. Schoua A. An English/Spanish test of decentering for the translation of questionnaires. Ann Arbor: University Microfilms International. 1985.

26. Harkness JA, Schoua-Glusberg A. Questionnaires in translation in cross-cultural survey equivalence. ZUMA Nachrichten Spezial No 3, 87–127. Mannheim, Germany. 1998.

27. Harkness JA, Pennell BE, Schoua-Glusberg A. Survey questionnaire translation and assessment. In: Presser S, Rothgeb JM, Couper MP, et al., eds, Methods for testing and evaluating survey questionnaires. New York: John Wiley and Sons. 2004.

28. Hinsdale M. Spanish translation issues related to the development of computer assisted interviewing instruments: National Household Survey on Drug Abuse. Paper presented at the Annual Meeting of the American Association for Public Opinion Research. Montreal, Quebec, Canada. 2001.

29. Ranjit N, Bakole A, Darroch J, Singh S. Contraceptive failure in the first two years of use: Differences across socioeconomic subgroups. Fam Plann Perspect 33(1): 19–27. 2001.

30. Henshaw S. Abortion incidence and services in the United States, 1995–96. Fam Plann Perspect 30(6):263–70 and 287. 1998.

31. Abma J, Sonenstein F. Sexual activity and contraceptive practices among teenagers in the United States, 1988 and 1995. National Center for Health Statistics. Vital Health Stat 23(21). 2001.

32. Frost J. Public or private providers? U.S. women's use of reproductive health services. Famy Plann Perspect 33(1):4–12. 2001.

33. Groves R., McGonagle K. A theory guided interviewer training protocol regarding survey participation. J Official Statistics 17(2):249–62. 2001.

34. Groves RM. Survey errors and survey costs. New York: John Wiley and Sons. 1989.

35. Groves R, Heeringa S. Responsive designs for household surveys. Paper presented at the Royal Statistical Society. London. 2004.

36. Groves RM, Singer E, Corning AD. Leverage-saliency theory of survey participation: Description and an illustration. Public Opinion Q 64:299–308. 1999.

37. Hansen MH, Hurwitz WN. The problem of nonresponse in sample surveys. J Am Stat Assoc 41:517–29. 1946.

38. American Association for Public Opinion Research. Standard definitions: Final dispositions of case codes and outcome rates for surveys. 2004. Available at: http://www.aapor.org. On-line ed 3.1. Revised 2005.

39. American Association for Public Opinion Research, interviewer falsification in survey research: Current best methods for prevention, detection and repair of its effects. Lexana, KS: AAPOR. 2004. Available at: www.aapor.org.

Appendix I. Glossary

ACASI—Audio computer-assisted self-interviewing, whereby the respondent uses a laptop computer to complete a questionnaire. The interviewer asks the respondent to use headphones; the respondent hears the questions through the headphones, or reads them on the computer screen. The respondent enters his or her answers to the questions using the laptop keyboard. The software directs the respondent to the next appropriate question. In the NSFG Cycle 6, the respondent performed these steps out of the sight of the interviewer, in an attempt to offer the respondent greater privacy.

Blaise®—A software system for computer-assisted questionnaires. Blaise® was developed by Statistics Netherlands and is licensed in the United States and Canada by Westat, Inc. When appropriately programmed, Blaise® can present the questions to the interviewer, route the respondent to the next appropriate question, record the answers entered by the interviewer, and check those answers for consistency with previous answers. Blaise® was used for Cycle 6 of the NSFG. An earlier version of Blaise® was used in Cycle 5. Further information is at http://www.westat.com/blaise.

CAPI—Computer-assisted personal interviewing, whereby an interviewer uses a laptop computer in the interview. The laptop displays question text for the interviewer to read, as well as providing any necessary instructions the interviewer may need. 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—The screener contact rate is the percentage of sample households where an interviewer talked with someone at the household. In 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 screener cooperation rate is the percentage of sample households who were contacted and granted a screener interview. The main interview cooperation rate is the percentage of sample persons contacted who granted a main interview.

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

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

Eligibility rate—The percentage of sample cases that are members of the target population. In NSFG Cycle 6 the eligibility rate is the percentage of households that contain a person 15–44 years of age.

Epsem—Equal probability selection method; a sample design that gives all sample units an equal chance of selection.

Life History Calendar—A hard-copy calendar covering the time periods referred to in various questions in an interview. The life history calendar is intended to help the respondent record key personal events, which are used as landmark events to cue memories of the dates of other events that are measured in the survey. In the NSFG Cycle 6 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 is eligible for the survey, a main interview is requested. In NSFG Cycle 6, if the household had one or more members 15–44 years of age, one such person was selected at random for the main interview request.

Paradata—Information describing the sample household, interactions with sample household members, or features of the interview situation. NSFG Cycle 6 used observations of characteristics of sample housing units to reduce the number of callbacks; used statements made by household members to diagnose concerns; used call record data to model response propensities; 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 NSFG Cycle 6 there were three phases: first, a quarter sample in which interviewers were permitted to visit households as many times as they thought was necessary (March-June 2002); second, the full sample with prespecified call rules (June 2002-January 2003), and third a subsample of nonrespondents from Phases 1–2, offered higher incentives by a subset of high productivity interviewers (February 2003).

Pretest—A survey done before a main survey. It is usually smaller than the main survey. Its purpose is to test questionnaires, procedures, equipment, and personnel before they are used in the larger main survey.

PSU—A primary sampling unit. The first stage selection unit in a multistage area probability sample. In NSFG Cycle 6 PSUs are counties or groups of counties in the United States; there were 121 PSUs selected into the NSFG sample.

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 in order to control the workflow of the interviewers. In responsive designs, early replicates are used to measure key cost and error features of a survey.

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

Sample line—A housing unit listed within a sample segment prior to the data collection phase of the NSFG, which was subsequently sampled for inclusion in the NSFG sample.

Screening interview—(Sometimes called a ''screener'') a (usually short) set of questions, asked of a household informant with the chief goal of determining whether anyone in the household is eligible for the chosen target population of the survey. In the NSFG Cycle 6, the screening interview consisted of a household roster, and collected age, race, and gender identification. Those households having one of more persons 15–44 years old were eligible for a main interview.

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 non-self-representing areas.

SurveyTrak—A software-based sample administration system. The system is used by interviewers on laptop computers to document their sample assignment, to organize the activities of their work day, to prompt them for appointments to be kept, to record results of each call attempt, to record observations of the sample housing unit, and in all other ways to keep track of their job duties.

Target population—The population to be described by estimates from the survey. In NSFG Cycle 6 the target population was the household population of the United States. Prisons, hospitals, dormitories, and other institutions are omitted. College students living in dormitories were interviewed but sampled through their parent/guardians' households.

Text CASI—(Text-only CASI) computer-assisted self-interviewing that does not allow the respondent to listen to an audio delivery of questions, but instead read the question text from the laptop's monitor.

WEBDOC—An electronic ''codebook'' for Cycle 6 of the NSFG, presenting data items, frequencies, code categories, and other survey documentation. The NSFG webdoc can be found at http://www.cdc.gov/nchs/about/major/nsfg/nsfgdoc.htm.

Appendix II. Outline of the Female and Male Questionnaires

Female Questionnaire Sections

Section	Main Topics
Section A	• Respondent demographic characteristics (age; DOB; marital/cohabitation status; race and Hispanic origin) • Household roster (age; sex; relationship of each member) • Introduction to Life History Calendar • Education (degrees; highest grade completed; date last attended)
Section B	• Childhood background and some information about her parents • Onset of menstruation (menarche) • Current pregnancy status • Number of pregnancies • Detailed pregnancy history (more details for pregnancies in last 5 years) • Care of nonbiological children • Relinquishment of biological children for adoption • Adoption (current plans and preferences; previous pursuit)
Section C	• Marital history and characteristics of each husband; more details for current husband • Details on current cohabiting partner, if there is one • Cohabitation history and selected details on each former cohabiting partner • Ever had sexual intercourse (asked if never married, never pregnant and never cohabited): NO: main reason why respondent has not had intercourse; YES: age and date of first intercourse • Details on first sexual partner (if not already discussed in the interview) • Date and age of first intercourse after menarche • Sex education (asked only of teens), including timing relative to first sex • Number of sexual partners (in lifetime; in past 12 months; before 1st marriage) • Recent (last 12 months) partner history; more details on current partners • Information on the respondent's last sexual partner (if no partners in last year)
Section D	• Sterilizing operations (respondent and husband/cohabiting partner) • Desire for sterilization reversal (only for tubal ligations and vasectomies) • Sterilizing operations among former husbands and cohabiting partners • Nonsurgical sterility and fertility problems (respondent and husband/cohabiting partner)
Section E	• Ever-use of contraceptive methods, including discontinued use and reasons for dissatisfaction with selected methods • Details on first method ever used (even if before first sexual intercourse) • Method use at first sexual intercourse • Months of intercourse for past 5 years or since first intercourse (later of 2 dates) • Contraceptive method history by month, for past 3 years or since first method used • Method used at first and last sex, with up to 3 partners in the last 12 months • Conditions surrounding respondent's pregnancies (including method use) • Wantedness of each pregnancy (by respondent and by father of pregnancy) • Happiness to be pregnant scale • Further details on circumstances surrounding pregnancies in last 3 years (including wantedness with that partner) • Current method use • Recent pill use (reasons; brand and type, consulting the Pill Chart) • Consistency of condom use (including frequency of sex in past 4 weeks)
Section F	• Use of medical services related to birth control and reproduction in the last 12 months and first time ever (services include receipt of: birth control method; checkup or medical test related to using birth control; counseling about birth control; counseling about getting sterilized; emergency contraception; information about emergency contraception; pregnancy test; abortion; Pap smear; pelvic exam; prenatal care; post-pregnancy care; testing or treatment for sexually transmitted disease (STD)) • Provider and payment information for each visit for these services in last 12 months (more detail if specific clinic is cited) • If clinic is regular source of medical care • For first visit ever for birth control services, asked date, what services, and provider • Ever visited a clinic

Female Questionnaire Sections—Con.

Section	Main Topics
Section G	• Desire for (wanting) a/another baby (respondent and husband/cohabiting partner) • Intentions to have a/another baby, asked individually or jointly, as appropriate
Section H	• Infertility services (to help get pregnant; to help prevent miscarriage) • Infertility diagnoses received, if ever pursued medical help • Vaginal douching • Health problems related to child-bearing (pelvic inflammatory disease; diabetes, both gestational and nongestational; ovarian cysts; uterine fibroids; endometriosis; problems with ovulation or menstruation) • Physical disabilities/limitations • HIV testing experience and knowledge about HIV
Section I	• Health insurance coverage in last 12 months • Current residence and residence as of April 1, 200 • Place of birth (date came to U.S. if born outside of the U.S.) • Religion and attendance of religious services, at age 14 and currently • Work status (respondent and husband/cohabiting partner) • Child care arrangements used (if any) in past 4 weeks for children under 13 • Attitudes: including relationships, sex, condom use, gender roles, and parenthood
Section J (ACASI)	• General health, including height and weight • Pregnancy history (numbers ending in live birth, abortion, or other outcomes) • Substance use (cigarettes; alcohol; marijuana; cocaine; crack; IV drugs) • Sexual intercourse with males (specific sexual behaviors she may have engaged in; condom use at last occurrence of vaginal, anal, or oral sex; condom use at last occurrence of any type of sex and reason; non-voluntary sex with males (asked only for respondents 18 or older); HIV/STD risk behaviors, including number of male partners) • Sex with females, including number of female partners • Sexual attraction and orientation • STD experience in last 12 months • Family income, sources of income and public assistance during 2001

Aids used in female questionnaire:
• Show Card booklet
• Question-by-Question help (part of the instrument)
• Life History Calendar
• Pill Chart

Male Questionnaire Sections

Section	Main Topics
Section A	• Respondent demographic characteristics (age; DOB; marital/cohabitation status; race and Hispanic origin) • Household roster (age; sex; relationship of each member) • Education (degrees; highest grade completed; date last attended) • Childhood background and some information about his parents • Numbers of marriages and cohabitations
Section B	• Sex education received (teen respondents only) • Sterilizing operations • Ever had sexual intercourse • Enumeration of (up to) three most recent female sexual partners
Section C	• Marital and cohabitation dates for current wife/partner • Sterilization and infertility (wife/partner) • Biological children with current wife/partner (more details if born in last five years) • Other children she had from previous relationships (more details if R adopted) • Nonbiological children with current wife/partner (more details if R adopted)
Section D	• Information on (up to) three most recent sexual partners (enumerated in B), including demographic data, dates and contraceptive use at 1st and most recent sex, and contraceptive use in last 12 months • Information on children with these partners (collected as above in C)
Section E	• Information on former wives and 1st (premarital) cohabiting partner (similar to information collected in C and D)
Section F	• Other biological children (information collected as above in C) • Other nonbiological children ever raised (more details if adopted, as above in C-E) • Pregnancies fathered in his lifetime that did not result in live birth (total number and numbers by outcome)
Section G	• Activities with the children living in his household • Activities with his biological and adopted children living elsewhere • Financial support of his biological and adopted children living elsewhere
Section H	• Desire for (Wanting) a/another baby (respondent and wife/cohabiting partner) • Intentions to have a/another baby, asked individually or jointly, as appropriate
Section I	• Usual source of health care • Health insurance coverage in last 12 months • Health services received in last 12 months (more details if under age 25) • Infertility services received • HIV testing experience and knowledge of HIV
Section J	• Current residence and residence as of April 1, 2000 • Place of birth (date came to U.S. if born outside of the U.S.) • Religion and attendance of religious services, at age 14 and currently • Work status (respondent and wife/cohabiting partner) • Attitudes: including relationships, sex, condom use, gender roles, and parenthood
Section K (ACASI)	• General health questions • Significant life events • Substance use (alcohol; marijuana; cocaine; crack; IV drugs) • Sexual intercourse with females (specific sexual behaviors he may have engaged in; condom use at last occurrence of vaginal, anal, or oral sex; condom use at last occurrence of any type of sex and reason; nonvoluntary sex with females (asked only for respondents 18 or older); HIV/STD risk behaviors, including number of female partners) • Sexual intercourse with males (specific sexual behaviors he may have engaged in; condom use at last occurrence of anal or oral sex; nonvoluntary sex with males (asked only for respondents 18 or older); HIV/STD risk behaviors, including number of male partners) • Sexual attraction and orientation • STD experience in last 12 months • Family income, sources of income and public assistance during 2001

Aids used in male questionnaire:
• Show Card booklet
• Q × Q help (part of the instrument)