

Supporting Statement B for Request for Clearance:

NATIONAL SURVEY OF FAMILY GROWTH, 2012-2015

OMB No. 0920-0314
(expires May 31, 2012)

March 6, 2012

Contact Information:

William D. Mosher, Ph.D., Statistician
Project Officer, National Survey of Family Growth
National Center for Health Statistics/CDC
3311 Toledo Road, Room 7318
Hyattsville, MD. 20782
301-458-4385
301-458-4034 (fax)
wmosher@cdc.gov

**Supporting Statement for Request for Clearance:
NATIONAL SURVEY OF FAMILY GROWTH,
Continuous Interviewing, 2012-2015**

PART B

B. Statistical Methods

NOTE: The sample design of the 2011-2015 NSFG is similar in most respects to the sample design of the 2006-2010 survey. The 2006-2010 survey is described in detail in the following two reports. The first is on the NSFG web site at www.cdc.gov/nchs/nsfg.htm. The second will be posted on the NSFG web site in Summer of 2012.

J Lepkowski et al. 2010. **The 2006-2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey.** Vital and Health Statistics, Series 2, No. 150. National Center for Health Statistics. June, 2010.

J. Lepkowski et al. **Innovation in Survey Research: Results of Fieldwork, Weighting, Imputation, and Variance Estimation in the 2006-2010 National Survey of Family Growth.** Vital and Health Statistics, Series 2. Publication expected Summer, 2012.

1. Respondent Universe and Sampling Methods

Summary—The National Survey of Family Growth (NSFG) is based on a national area probability sample. To control costs, the sample is being drawn from a nationally representative sample of only 35 Primary Sampling Units (PSUs) each year, but the PSU's rotate each year, so that a national sample of 117 PSU's will be used in 4 years. The data will be collected annually and continuously. Each year, about 14,000 households will be contacted, in order to yield the required 5,000 interviews annually. Each year of data is an independent national sample, but the desired sample size and precision will be attained after 4 years of interviewing (Sept 2011-Sept 2015).

Target Population. The target population of the National Survey of Family Growth is the household population 15-44 years of age. It excludes current residents of military bases and institutions (e.g., long-term hospitals, jails, prisons). College students temporarily away from their homes at college are included by sampling them at their home address; they can be interviewed either at home or at college.

Details of the Sample Design

- (1) 117 Primary Sampling Units (counties, or groups of adjacent counties) are selected at random from the entire set of more than 3,100 counties in the US, including Alaska and Hawaii. PSUs are selected with probability proportionate to population size—that is, counties and groups of counties comprising metropolitan statistical areas (MSA's) with large populations have a larger chance of selection, and the three MSA's with the largest populations are always included.
- (2) Within each of those 117 PSU's, smaller areas called sample segments are selected, again at random. A segment is a geographical area (like a group of blocks in cities or an area bounded by roads in a rural area). It can contain as few as 50 structures in a rural area or several hundred in a densely settled urban area.
- (3) Trained staff are sent out to prepare a list of addresses in the segment. (In urban areas, the listers are verifying and correcting a commercially purchased address list; in rural areas, they are listing from scratch.)
- (4) Once the addresses are listed, a sample of the listed addresses is selected (again, by chance) for the study.
- (5) After an advance letter is sent to each selected household informing them about the study (**Attachment G1**), a trained survey interviewer visits the household, to collect a household roster (or screener), in order to see if someone 15-44 years of age lives there. If more than one person is 15-44 and eligible, then one person is selected at random for the interview.
- (6) The data collection activities of the National Survey of Family Growth are continuous, but during each year of the survey, a randomly selected subset of the PSUs is used, so each year

is a probability sample of the US household population, albeit smaller than the full sample. This annual sample consists of about 35 primary sampling areas throughout the country.

- (7) At the end of each year, the PSUs are dropped from the sample and new PSU's are included. At the end of four years, the cumulative sample contains the full set of about 117 PSUs.

The *rotating* feature of the PSUs permits a cost efficiency of ongoing sampling and data collection operations by using the field interviewing staff and funding in an optimal manner. It further offers at any single year a full national sample for the study, albeit with standard errors of estimates larger than those of the 4-year cumulative sample.

Group quarters with special living arrangements, such as dormitories, institutions, convents, or institutional group homes (for convicts, the frail elderly, or the developmentally disabled, e.g.) may be listed but will not be selected for interviewing, because they are outside the scope of a sample of the household population. Dormitory residents who otherwise live with their parents will be sampled at their parents' homes. Members of the active duty military who live in civilian housing (not on military bases) will be eligible for the sample. The NSFG is a personal visit survey. Telephone contacts are permitted only to arrange appointments for interviews after the screener has been conducted, and for 5-minute verification interviews (**Attachment K**) to ensure that the respondent was interviewed.

2. Procedures for the Collection of Information

The sample size targets are as follows:

(see next page)

**Sample Size Targets for NSFG Continuous Interviewing
with 2002 (Cycle 6) and 2006-10 sample sizes shown for comparison**

	<u>Cycle 6</u> 2002	<u>4-year</u> <u>Continuous</u> 2006-2010	<u>4-year</u> <u>Continuous</u> 2011-2015
TOTAL	12,571	22,682	20,000
15-19	2,271	4,662	4,000
20-44	10,300	18,020	16,000
Male	4,928	10,403	9,000
Female	7,643	12,279	11,000
Hispanic	2,712	5,132	4,000
Black	2,460	4,389	4,000
White & other	7,399	13,161	12,000

The sample sizes cumulated in 2011-15 will allow estimates for small but important groups such as Hispanic male teenagers, couples who have adopted children, fathers who do not live with their children, infertile women 35 years of age and older, gay and lesbian populations, and those who are at risk of HIV because of their sexual behavior.

The current contractor for the NSFG is the University of Michigan's Institute for Social Research (ISR; Mick Couper, Project Director, and Nicole Kirgis, Field Director). Under the supervision and monitoring of NCHS, ISR recruits and trains the interviewers for the NSFG and carries out the fieldwork. The main steps in the fieldwork are described below.

Main steps in field work:

- (1) Before contacting households, the contractor will send an advance letter and pamphlet to all eligible households. These explain who is sponsoring the survey, who is conducting it, why it is being done, and the voluntary and confidential nature of the survey. Spanish versions of the questionnaires, the advance letter, and other introductory materials are also prepared, as in past Cycles of the survey. (The NSFG has

had a Spanish version of the questionnaire since 1973.) The letters and informed

consent materials are shown in **Attachments G1-G3**.

- (2) When the housing unit is found to be occupied and the interviewer finds someone (18 or older) at home, the screener interview (**Attachment G4**) is conducted. The purpose of the screener is to list the persons living in the household and their ages, and if one or more are 15-44 years of age, to select one. Age and gender are collected in the screener because teenagers and women are selected at somewhat higher rates than adults and men.
- (3) When a person 15-17 years of age is selected for the sample, signed parental consent is obtained before the interviewer introduces the survey to the teenager. A parent letter and consent form will be used to explain the survey to the minor's mother, father, or guardian, and ask for their written consent.

If either the parent or the minor Respondent refuses to give consent, the case is treated as a refusal. If the parent gives consent, then the interviewer introduces the survey to the 15-17 year old, and asks the teenager for his or her signed assent. If the teen is willing, he or she signs the "Minor Assent" form (**Attachment G3**), and proceeds to the interview.

Emancipated minors--15-17 year-olds who are married or cohabiting and living away from their parents are rare in a sample of this size. Emancipated minors have been excluded from the continuous NSFG because the number of emancipated minors selected for the NSFG is so small that excluding this group is unlikely to have any noticeable impact on estimates. Using current IRB rules, however, including them would require special procedures that are too complex and too costly for the NSFG.

- (4) If the Respondent is 18 years of age or older, the interviewer gives the Respondent an Adult Consent Form (**Attachment G3**), which explains the survey and requests signed consent. If the Respondent agrees to do the survey but refuses to sign the form, the interviewer can offer to begin the interview, and ask for a signature at the end of the interview, or sign for the respondent.
- (5) The interviewer gives the respondent \$40 as a “token of appreciation.” The respondent can keep the incentive even if he or she does not finish the interview. (Break-offs are rare in this survey—less than 1 percent.)
- (6) Then the interview is conducted (**Attachments H-1 and I-1**), using a laptop computer. This use of the computer makes the interviewer's job easier, reduces interviewer errors, protects confidentiality, and produces higher quality data.
- (7) Finally, at the end of the interviewer-administered interview, the interviewer gives the respondent a pair of headphones and the notebook computer, and shows the respondent how to make simple entries on the computer. The respondent then completes a 10-15 minute Audio Computer-Assisted Self-Interview (**Audio CASI**). The interviewer cannot see or hear what questions the respondent is being asked over the headphones, and cannot see or hear the answers that the respondent enters into the computer. Moreover, no one in the household can hear or see either the questions or the answers. (The screen can be blanked with one keystroke.) This increased privacy has been found to increase the reporting of sensitive behaviors.

While the respondent is filling out the Audio CASI part of the interview, the interviewer fills out the Interview Observation Form (**Attachment J**), which formalizes some field notes that have been collected in less structured form since the 1973 NSFG, on where the interview was done, whether there were interruptions during the interview, and the interviewer's assessment of the quality of the data. (The Interview Observation Form is

filled out by the interviewer; no questions are asked of the respondent.)

- (8) At the end of the Audio CASI section, the interviewer turns off the computer, thanks the respondent, and leaves. The interviewer cannot back up and see the respondent's answers, because the Audio CASI system is locked by the respondent when he or she is done.

Quality control

Computer-assisted interviewing improves data quality in several ways:

- (a) Interviewer errors are reduced because interviewers do not have to follow complex routing instructions; the computer does it for them. Interviewer errors in following skip patterns were a principal cause of missing data in paper and pencil interviewing.
- (b) Respondent errors are also reduced with CAPI interviewing. The contract requires that selected consistency checks be programmed into the questionnaire so that inconsistent answers can be corrected or explained while the interview is still in progress. We continue to work on identifying and resolving logical inconsistencies earlier and more efficiently than in the past, to improve data quality and expedite data release.
- (c) Coding and coding errors are also reduced using CAPI interviewing, and this makes it possible to prepare the data for analysis faster and more accurately. In Continuous Interviewing, earlier cases (e.g., year 1) are being used to discover and correct errors before they affect later cases (e.g., year 2).
- (d) The "Verification" interview is a quality control procedure in which a random sample of both respondents and non-respondents will be contacted (usually by telephone) after the interview to verify that the interview was conducted with the appropriate sample person. **(Attachment K)**
- (e) Editing -- Additional computer editing of the data will be performed by the Contractor in the home office after the interviews are complete. NCHS is also performing checks of

the quality of the data files, as it has in past cycles.

- (f) Imputation -- A few hundred of the most frequently used variables (called "Recodes") are imputed when missing. On most of these items, missing data was less than 1 percent. The imputation procedure is described further in:

Vital and Health Statistics, Series 2, No. 142, "National Survey of Family Growth Cycle 6: Sample design, weighting, imputation, and variance Estimation," July 2006; see the NSFG web site at www.cdc.gov/nchs/nsfg.htm).

Two basic types of imputation were used for about 600 variables (out of about 6,000 variables on the data file):

- regression model-based imputation (used for most variables)
- logical imputation (for a few variables with only a handful of missing cases).

The large majority of imputations are being done by multiple regression imputation using the University of Michigan's Imputation and Variance Estimation software, which is called "IVEWARE." As in previous cycles, the public use data files have imputation "flags"—variables that show that a value was imputed--so that data users can assess for themselves whether imputation affects the estimates. Imputation rarely affects estimates in the NSFG because the levels of missing data are generally very low.

- (g) Estimation -- Estimation refers to the process of producing weighted numbers and percentages for the population from sample data. For each case, a weight is generated which estimates the number of persons in the population that each sampled person represents. For example, if a woman represents 5,000 women in the population, her weight is 5,000. The weight for each respondent is created in 4 basic steps:

- inflation by the reciprocal of the probability of selection,
- adjustment for nonresponse within age, sex, and race categories,
- post-stratification to independent control totals provided by the Census Bureau, and
- trimming of a small number of extreme weights.

Probabilities of selection vary because black, Hispanic, and teenage respondents are oversampled, and because non-respondents to the survey are sub-sampled for the “double sample” in the non-response follow-up (the last phase of data collection). Adjustments for non-response are made by multivariate (logistic regression) methods. Post-stratification to control totals is done within cells defined by race and origin, age, and sex.

Variances are being estimated using a Taylor Series linearization approach similar to that used in the 2002 NSFG and described in **Series 2, No.142** (available on the NSFG web site at www.cdc.gov/nchs/nsfg.htm, under “Publications and Information Products.”). Codes were generated that allow data users to compute variances using Taylor Series linearization, Balanced Half-Sample Replication, or Jackknife replication methods. A similar procedure will be used to produce the 2011-15 data file.

3. Methods to Maximize Response Rates and Deal with Non-response

As discussed above, we use Advance Letters, highly trained interviewers, a web site, 800 numbers at both the University of Michigan and at NCHS, customized follow-up letters to address particular concerns, and special interviewer training on non-response, to encourage cooperation with the survey, and active survey management using daily paradata to allocate interviewer effort. Our principal guidance in dealing with non-response is our experience in the 2002 and 2006-2010 NSFG, as documented in:

R Groves, G Benson, W Mosher, et al. 2005. **Design and Operation of Cycle 6 of the National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 42**, August, 2005. National Center for Health Statistics, Hyattsville, MD. Available at: <http://www.cdc.gov/nchs/nsfg.htm>.

R Groves and SG Heeringa. 2006. **Responsive Design for Household Surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society A169, Part 3: 439-457**, April, 2006.

R. Groves, W Mosher, et al. 2009. **Planning and Development of the Continuous National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 48.** National Center for Health Statistics. September 2009.

J Lepkowski et al. **The 2006-2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey. Vital and Health Statistics, Series 2, No. 150.** National Center for Health Statistics. June, 2010.

J. Lepkowski et al. **Innovation in Survey Research: Results of Fieldwork, Weighting, Imputation, and Variance Estimation in the 2006-2010 National Survey of Family Growth. Vital and Health Statistics, Series 2.** Publication expected in Summer 2012.

Procedures are listed separately for non-contacts, and for refusals. For non-contacts, the following procedures are used:

- (a) listers of sample segments document units that have access impediments (e.g., locked apartment buildings, or security guards at a community entrance gate). Interviewers will schedule calls on such cases earlier in the field period than others,
- (b) observations are made by the interviewer regarding best times to reach the sample household, and
- (c) multiple calls are made on sample units, at different times of the day and different days of the week.

For refusals, interviewers are trained to avert refusals by understanding and learning to respond to the concerns that potential respondents express. Letters on NCHS letterhead, signed by the NCHS Director, are used for all sample households to communicate the scientific goals and practical usefulness of the research and to legitimate the visit of the interviewer. Letters to local community police are also used in some areas to announce the presence of interviewers in the area. Interviewers are in ongoing contact with their supervisors, allowing interviewers to seek guidance on individual problems they encounter. Throughout this process interviewers are explicitly instructed to treat the sample person's concerns as legitimate questions that deserve thoughtful answers. Our approach is to answer respondents' questions. Emphatic or "hard" refusals are accepted as final.

Guidance to interviewers in continuous interviewing is based on the research and experience cited above, and on extensive paradata—data about the fieldwork—collected and recorded by interviewers and other field staff. These data are summarized using logistic regression equations into a total propensity to respond for an entire segment. These data (and case-specific observations entered into the contractor's sample management system) can be used to guide further actions on individual cases. (*Paradata are discussed in the reports cited on pages 10 and 11.*)

Incentives.—Given that even the good survey practices described above are unlikely to attain an 80% response rate with the budget available to the NSFG, in Section A9, we requested OMB clearance to continue to use a \$40 cash incentive in 2012-2015. Previous research (cited below and in **Attachment C**) suggests that, for long, sensitive, in-person surveys, incentives do help raise response rates and help to control fieldwork costs when standard good survey practice is not enough.

Incentives at the \$40 level appear to be especially effective among minorities, teenagers, and low-income people. And since low-income people have (for example) very different patterns of contraceptive use, unintended pregnancy, and marriage and divorce than high-income people, our results would be biased without the use of incentives. That observation is consistent with the NSFG's experience in the 2002 NSFG and in 2006-2010. Given that interviewer labor costs about \$25 an hour (including indirect costs and supervisor time), this \$40 amount quickly pays for itself, because it saves interviewers time.

At the same time, we have also found (see **Attachment C**) that incentives at the \$80 level (given to just 6% of completed interviews in 2006-2010) are necessary to increase participation from busy, high-income, married, well-educated respondents. This group also has some distinctive behavioral patterns that would be under-represented if we did not use the follow-up to bring them into the sample.

Examples of the literature that guides our use of incentives is shown below:

E. Singer, "The Use of Incentives to Reduce Nonresponse in Household Surveys," pages 163-178 in R Groves et al (editors), Survey Nonresponse. Wiley, 2002.

Kulka R. "The Use of Incentives to Survey 'Hard to Reach' Respondents," pages 256-287, In: Federal Committee on Statistical Methodology, Statistical Policy Working Paper No. 23, Volume 2.

Robert M. Groves, Mick P. Couper, Stanley Presser, Eleanor Singer, Roger Tourangeau, Giorgina Piani Acosta, and Lindsay Nelson. 2006. "Experiments in Producing Nonresponse Bias," Public Opinion Quarterly, 2006; 70: 720 - 736.

Davern, Michael; Todd H. Rockwood, Randy Sherrod, and Stephen Campbell. 2003. Prepaid Monetary Incentives and Data Quality in Face-to-Face Interviews: Data from the 1996 Survey of Income and Program Participation Incentive Experiment. Public Opinion Quarterly, Spring 2003; 67: 139 - 147.

Nonresponse Bias Studies Planned

Attachment N describes our approach to measuring and managing nonresponse bias. Procedures to measure and reduce nonresponse bias are built into the daily paradata monitoring of the study. NSFG has the following data resources to warn us of possible nonresponse bias and allow us to act to reduce it:

- 1) The NSFG's paradata includes observations from listers and interviewers. Their observations include variables such as whether the building is locked or access is blocked by other barriers, whether the household includes children, the marital status of the screener respondent, and others that are correlated with non-response or NSFG outcome variables.
- 2) Key statistics (percent married, percent who have had a child, etc) are tracked to see if they change when calling effort is increased;

- 3) We monitor daily the response rates of 12 age-race-gender groups that are strongly correlated with many NSFG estimates (e.g., Hispanic males 20-44; black females 15-19). If rates are strongly unequal, that inequality could cause biased estimates. Effort can then be increased on groups with lagging response rates. If more effort is required for certain groups, effort can be directed toward those groups during fieldwork. Such interventions can be randomized, so that their effects can be measured.
- 4) A two-phase sampling scheme is used. At the end of 10 weeks of fieldwork, a probability sample of nonrespondents is selected. Incentives are increased for the selected cases, and different fieldwork techniques are used. Response rates and sample composition can be compared before and after “phase two” of fieldwork.
- 5) Alternative post-survey adjustments for nonresponse can be compared.

A more complete description of these activities appears in Attachment N.

4. Tests of Procedures or Methods

The first several weeks of interviewing in 2011 served as the pretest for the new cycle of continuous interviewing. These interviews have gone well. We do not believe that additional pretesting will be required until 2015, when new questions will again be introduced.

However, the first several weeks of interviewing did reveal that the female questionnaire (**Attachment H-2**) was about 9-10 minutes longer than its stated length of 80 minutes. As a result, after consultations with other interested programs, we are deleting certain questions from the questionnaires (listed in **Attachment B1**, and shown in context in **Attachments H-2 and I-2**).

We estimate that they will restore the female questionnaire to its stated length of 80

minutes, and reduce the male questionnaire to 57-58 minutes.

5. Statistical Consultants

The statistical consultant (on sample design, variance estimation, and statistical methods) for NCHS is:

Van L. Parsons, Ph.D.
Mathematical Statistician
NCHS Office of Research and Methodology
301-458-4421 e-mail: vparsons@cdc.gov

The sample selection and data collection are supervised for NCHS by:

William D. Mosher, Ph.D.
Project Officer, NSFG
NCHS, Room 7421
3311 Toledo Road
Hyattsville, MD 20782
301-458-4385 e-mail: wmosher@cdc.gov

Sample selection and data collection are supervised for the contractor by:

Mick Couper, Ph.D., Project Director, NSFG, and
Associate Director, Survey Research Center,
University of Michigan
426 Thompson St, Ann Arbor, MI 48104
734-647-3577 mcouper@isr.umich.edu

James Wagner, Ph.D.
Senior Mathematical Statistician, NSFG
Institute for Social Research
University of Michigan
426 Thompson Street, Ann Arbor, MI 48104
734-647-5600 jameswag@isr.umich.edu

The person responsible for the analysis of the survey is:

William D. Mosher, Ph.D, Project Officer for NCHS: Phone: 301-458-4385
e-mail: wmosher@cdc.gov

LIST OF ATTACHMENTS

A. Authorizing legislation

- A1. NSFG Authorizing Legislation
- A2. Office of Population Affairs Authorizing Legislation
- A3. NICHD Authorizing legislation
- A4. Children's Bureau (ACF) Authorizing Legislation
- A5. OASPE (Office of the Assistant Secretary for Planning and Evaluation)
Division of HIV/AIDS Prevention, CDC; and
Division of Sexually Transmitted Disease Prevention, CDC
- A6. Office of Planning, Research, and Evaluation (OPRE), ACF
- A7. Division of Cancer Prevention and Control, CDC
- A8. Division of Birth Defects and Developmental Disabilities, CDC

B1. List of Questions Deleted from the NSFG

B2. Justifications for Sensitive Questions in the Self-administered (ACASI) part of the Survey

C. A Review of the Use of Incentives in the NSFG

D. Partial list of publications from the Survey

- D1. List of publications from the 2002 NSFG
- D2. List of publications from the 2006-2010 NSFG.

E. Memoranda from other offices and agencies on their use of the NSFG

- E1. NCHS Public Affairs Officer
- E2. Healthy People 2010 Health Objectives on Family Planning, HIV, STDs

F. Consultation outside the agency:

- F1. Agenda for the October, 2008 Research Conference on the NSFG.
- F2. Agenda for the November, 2008 Meeting of the NSFG advisory workshop
- F3. Report of the NSFG Review Group for the NCHS Board of Scientific Counselors, April, 2010
- F4. 60-Day Notice for the National Survey of Family Growth, 2011
- F5. Report on our trip to CDC/Atlanta, November 15-17, 2010
- F6. Response to public comment on 60-day notice

G. Respondent Materials for the NSFG in 2009-2012

- G1. Respondent Letters
- G2. Brochures and Letter of Authorization

G3. Consent Forms

G4. Screener Questionnaire

H. FEMALE Questionnaire

H1. Female Questionnaire, 2012-2015 (clean, unmarked)

H2. Female Questionnaire, 2011 (with deleted questions shown in red)

I. MALE Questionnaire

I-1. Male Questionnaire, 2012-2015 (clean, unmarked)

I-2. Male Questionnaire, 2011 (with deleted questions shown in red)

J. Interview Observation Form (filled out by the Interviewer)

K. Verification Questionnaire

L. (Not Used)

M. IRB Approval Forms for the NSFG in 2011

N. Non-Response Bias Analyses for the continuous NSFG