SF-83-1 SUPPORTING STATEMENT

for

2008
National Survey of College Graduates

CONTENTS

Section **Page**

A. JUSTIFICATION 1
1. Necessity for Information Collection 2
2. Uses of Information 4
3. Consideration of Using Improved Technology 5
4. Efforts to Identify Duplication 6
5. Efforts to Minimize Burden on Small Business 6
6. Consequences of Less Frequent Data Collection 6
7. Special Circumstances 7
8. Federal Register Announcement and Consultation Outside the Agency 7
9. Payment or Gifts to Respondents 10
10. Assurance of Confidentiality 10
11. Justification for Sensitive Questions 11
12. Estimate of Respondent Burden 11
13. Cost Burden to Respondents 12
14. Cost Burden to Federal Government 12
15. Reason for Change in Burden 12
16. Schedule for Information Collection and Publication 12
17. Display of OMB Expiration Date 13
18. Exception to the Certification Statement 13

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS 14
1. Respondent Universe And Sampling Methods 14
2. Statistical Procedures 14
3. Methods to Maximize Response 15
4. Testing of Procedures 18
5. Contacts for Statistical Aspects of Data Collection 22

APPENDIX A: NSF Act of 1950
APPENDIX B: First Federal Register Announcement
APPENDIX C: Proposed 2008 NSCG Sampling Tallies
APPENDIX D: NSCG Questionnaire Changes
APPENDIX E: Survey Mailing Materials
APPENDIX F: Draft 2008 NSCG Questionnaire
APPENDIX G: 2008 NSCG Incentive Conditioning Experiment

2008 NATIONAL SURVEY OF COLLEGE GRADUATES
SUPPORTING STATEMENT

A. JUSTIFICATION

This request is for a three-year reinstatement, with change, of the previously approved OMB clearance for the National Survey of College Graduates (NSCG). The NSCG was last conducted in 2006. The OMB clearance for the 2006 NSCG expires February 28, 2009.

The NSCG is one of three principal surveys that provide data for the NSF's Scientists and Engineers Statistical Data System (SESTAT). The purpose of the SESTAT database is to provide information on the entire U.S. population of scientists and engineers with at least a bachelor's degree. SESTAT is produced by combining data from the Survey of Doctorate Recipients (SDR; representing persons in the general U.S. population who have earned a doctorate in science, engineering, or health (SEH) from a U.S. institution), the National Survey of Recent College Graduates (NSRCG; representing persons with a recently earned bachelor's or master's degree in SEH from a U.S. institution) and the NSCG (representing all individuals in the U.S. at the time of the decennial census with a bachelor's or higher degree in a science or engineering (S&E) or S&E-related degree, or those who had a bachelor's or higher degree in some other field, but had an S&E or S&E-related occupation, including individuals who received degrees only from foreign institutions). The NSCG population is primarily drawn from eligible individuals from the decennial census.

The SESTAT integrated database derived from these surveys represents the demographic, educational, and employment characteristics of college-educated scientists and engineers in the United States. All three of these surveys are usually conducted every two years. The primary purpose of the NSCG is to provide information on the U.S stock of scientists and engineers early in each decade. The panel portion of the SDR also provides information on the stock, while the new sample in the SDR and the entire NSRCG provide important data on the new graduates with SEH degrees entering the labor force during each decade. The NSCG constitutes the bulk of the records in the SESTAT database; accounting for approximately 60% of the records in the SESTAT system and slightly over 90% of the population estimate for 2006.

The SESTAT integrated database is the only available source that provides detailed information to support a wide variety of policy and research analyses on SEH personnel. To provide complete representation of U.S. SEH at all degree levels, SESTAT was designed as a unified database that integrates information from all three component surveys. The system of surveys, created for the 1993 survey cycle and developed throughout the 1990s, is closely based on the recommendations of the National Research Council's Committee on National Statistics (CNSTAT) report to NSF[1]. That report recommended a data collection design based on three surveys, of which one (the NSCG) would be linked to the decennial Census.

[1] National Research Council. Committee on National Statistics. (1989). Surveying the Nation's Scientists and Engineers: A Data System for the 1990s. Washington: National Academy Press.

The 1993 NSCG served as the baseline survey of the decade of the 1990s and interviewed a large number of individuals in order to identify those who were scientists and engineers, based on their education or occupation. Because the decennial census does not collect information on field of degree, the sample was of all persons with a bachelor's degree or higher in order to find persons educated in all fields of science and engineering, and persons educated at foreign universities, as well as persons working in science and engineering occupations. The 1995, 1997, and 1999 NSCG panel studies followed through the decade individuals identified in the 1993 survey as having an S&E degree and/or an S&E occupation. The 2003 NSCG, like the 1993 survey, was the "baseline" survey for this decade and was used as a screener survey to identify those who are scientists and engineers based on their education or occupation. The respondents to the 2006 NSCG and new science, engineering and health bachelor's and master's degree recipients that responded to the 2006 NSRCG will be contacted in the 2008 NSCG panel survey.

NSF incorporated the lessons learned from 1990s into the sample design and weighting for the 2003 NSCG survey. It also provided NSF with the opportunity to evaluate all aspects of the design and content of all the surveys contributing to the SESTAT database (described in section A.8). Changes recommended in this review process were incorporated in the 2003 NSCG to the extent possible and will continue in the 2008 NSCG. 2008 questionnaire content will not include any new questions but will include a module of questions that were used in the past. The changes made to the questionnaire are described with the reasons for these changes in Section B.4. The 2008 NSCG draft questionnaire is presented as Appendix E.

1. NECESSITY FOR INFORMATION COLLECTION

The National Science Foundation Act of 1950, as amended by Title 42, United States Code, Section 1862 requires the National Science Foundation to:

> "Provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources and to provide a source of information for policy formulation by other agencies of the Federal Government..." (See Appendix A)

In meeting its responsibilities under the NSF Act, the Foundation relied on the National Register of Scientific and Technical Personnel from 1954 through 1970 to provide names, location, and characteristics of U.S. scientists and engineers. Acting in response to a Fiscal Year 1970 request of the House of Representatives Committee on Science and Astronautics (see U.S. Congress, House of Representatives, 91st Congress, 1st Session, *Report No. 91-288*), the Foundation, in cooperation with the Office of Management and Budget and eight other agencies, undertook a study of alternative methods of acquiring personnel data on individual scientists and engineers.

The President's budget for Fiscal Year 1972, as submitted to the Congress, recommended the "discontinuation of the National Register of Scientific and Technical Personnel in its present form" and that funds be appropriated "to allow for the development of alternative mechanisms for obtaining required information on scientists and engineers." The House of Representatives Committee on Science and Astronautics in its report on Authorizations for Fiscal Year 1972 states that "...it has no objection to this recommendation...." (See U.S. Congress, House of Representatives, 92nd Congress, and 1st Session, *Report No. 92-204*).

Subsequently, the NSF established and continues to maintain the SESTAT system of surveys, the successor to the Scientific and Technical Personnel Data System of the 1980s, which was the successor to the National Register. The Science and Technology Equal Opportunities Act of 1980 directs NSF to provide to Congress and the Executive Branch an "accounting and comparison by sex, race, and ethnic group and by discipline, of the participation of women and men in scientific and engineering positions." The SESTAT database, of which NSCG is the large majority of records, provides much of the information to meet this mandate.

The longitudinal data from the NSCG provides valuable information on training, careers and educational development of the Nation's highly educated science and engineering population. These data enable government agencies to assess the scientific and engineering resources available in the U.S. to business, industry, and academia, and to provide a basis for the formulation of the Nation's science and engineering policies. Educational institutions use NSCG data in establishing and modifying scientific and technical curricula, while various industries use the information to develop recruitment and remuneration policies.

The NSF uses the information to prepare congressionally mandated biennial reports such as *Women, Minorities and Persons with Disabilities in Science and Engineering* and *Science and Engineering Indicators.* These reports enable NSF to fulfill the legislative requirement to act as a clearinghouse for current information on the S&E workforce.

The Committee for Equal Opportunity in Science and Engineering (CEOSE), an advisory committee to the NSF and other government agencies, established under 42 U.S.C. §1885c, has been charged by the U.S. Congress with advising NSF in assuring that all individuals are empowered and enabled to participate fully in science, mathematics, engineering and technology. Every two years CEOSE prepares a congressionally mandated report that makes extensive use of the SESTAT data to highlight key areas of concerns relating to students, educators and technical professionals.

Congress also created the Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development (P.L.105-255, October 1998). In this legislation, the Commission was mandated to analyze and describe the current status of women, underrepresented minorities, and persons with disabilities in the science, engineering, and technology pipeline from early classroom education through their professional lives in industry, academe, and government. The final report, *Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering and Technology* (September 2000), made extensive use of the SESTAT data to answer critical questions on the status of these groups in the workforce.

The importance of information on the scientific and technical workforce to inform public policy can be seen in discussions of the National Science Board's Task Force on National Workforce Policies for Science and Engineering. The taskforce relied heavily on SESTAT data to inform its deliberations about the S&E workforce and SESTAT data were an integral part of the taskforce's final report. (See http://nsf.gov/nsb/documents/2003/nsb0369.)

2. USES OF INFORMATION

Researchers, policymakers and other users of the data use information from the SESTAT database to answer questions about the number, employment, education, and characteristics of the S&E workforce. Because it provides up-to-date and nationally representative data, researchers and policymakers use the database to address questions on topics such as the role of foreign-born or foreign-degreed scientists and engineers, the transition from higher education to the workforce, the role and importance of postdocs, diversity in both education and employment, the implications of an aging cohort of scientists and engineers as baby boomers reach retirement age, and information on long-term trends in the S&E workforce.

Data from NSF's SESTAT component surveys are used in policy discussions of the executive and legislative branches of Government, the National Science Board, NSF management, the National Academy of Sciences, professional associations, and other private and public organizations. Some recent specific examples of the use of the SESTAT data are: the Urban Institute used the SESTAT data in the evaluation of the NSF's Louis Stokes Alliance for Minority Participation Program; the Society of Women Engineers referenced SESTAT data in their 2006 Literature Review and provides links to the SESTAT database on their website; and Ph.D. students use the SESTAT workforce data in dissertations.

Data Dissemination and Access

The NSF makes the data from the SESTAT system of surveys available through published reports, the SESTAT on-line data system, public use files and restricted licenses. The 1993 and 2003 NSCG data are available as a public-use file. The NSCG panel data from all the 1990s and 2000s cycles are also available as a component of the SESTAT data base for each survey year (1993, 1995, 1997, 1999, 2003 and 2006) and are available as SESTAT public-use files.

The SESTAT data were used extensively in the latest versions of the congressionally mandated biennial reports *Science and Engineering Indicators, 2008* and *Women, Minorities and Persons with Disabilities in Science and Engineering, 2007.*

NSF also used the NSCG and SESTAT integrated data in recent reports such as:

- *Unemployment Rate of U.S. Scientists and Engineers Drops to Record Low 2.5% in 2006*, March 2008
- *Why Did They Come to the United States? A Profile of Immigrant Scientists and Engineers*, June 2007
- *What Do People Do After Earning an Science and Engineering Bachelor's Degree?* June 2006
- *2003 College Graduates in the U.S. Workforce: A Profile*, December 2005

All NSF Publications can be accessed on the SRS website at http://www.nsf.gov/statistics.

To provide better accessibility to information for policy makers and researchers, NSF provides the SESTAT integrated database and the NSCG data on the World Wide Web. The SESTAT on-

line system allows Internet users to create customized data tabulations with a user-specified subject area. The SESTAT Home Page can be accessed at http://www.nsf.gov/statistics/sestat.

Results from the SESTAT integrated database and NSCG data are routinely presented at the conferences and professional meetings, such as the annual meeting of the Association for Institutional Research or the American Educational Research Association.

Since 2005, NSF has distributed over 200 files of the almost decade-old 1993 public-use NSCG data set and close to 400 files of the 1993 SESTAT integrated database public-use version to researchers in government, academia, and professional societies. In spite of the age of the data, the 1993 NSCG data continue to be heavily used because it is the only data set analysts can use to compare the S&E workforce to the general population of college degree holders in the U.S. Besides capturing people with degrees earned at U.S. institutions, the NSCG also includes college degree holders who earned their degrees outside of the United States and who were residing here at the time of the last census. The 2003 NSCG public-use data also has been widely distributed to over 380 users. Over 18 licensed users have accessed to the 2003 SESTAT integrated database micro data file under the licensing agreement with SRS. As previously noted, over 60% of the records in this file come from the NSCG.

Some of the research from the NSCG and SESTAT licensees resulted in papers such as:

- *In Search of the Glass Ceiling: Cohort Effects on Women's Wage,* University of California Santa Barbara, 2005
- *The Effect of IT on the Publication Gap Between Women and Men in Academia,* University of Missouri St. Louis, 2005
- *Entrepreneurship and Advanced Technical Knowledge*, State University of New York Buffalo, 2005
- *Returns to Graduate and Professional Education: The Roles of Mathematical and Verbal Skills by Major*, Iowa State University, 2004

3. CONSIDERATION OF USING IMPROVED TECHNOLOGY

The NSCG data will be collected by the U.S. Census Bureau, under interagency agreement, using a multi-mode approach; that is, a questionnaire will be mailed to sample persons and the nonrespondents will be followed up using computer-assisted telephone interviewing (CATI). Because the sample contact information will be at least 2-½ years old by the time the survey is conducted, extra effort will have to be spent to locate respondents. To do this in the most efficient way, the NSCG will employ nonintrusive locating procedures to find valid mailing addresses for cases that are identified as nonmailable after the sample is sent through automated software to check against updates to the National Change of Address (NCOA) database. These nonintrusive procedures include the use of Internet search engines, and name and address locating software such as FastData and InfoUSA. Additionally, the Census Bureau has developed an electronic locating system to improve the efficiency of the locating operation.

The 2008 NSCG will use Census' state-of-the-art keying system, developed in the Visual Basics software, to capture mail questionnaire data, which should decrease the time necessary

for this operation and increase accuracy. The telephone interviewing phase will utilize a variety of improved technologies. Interviews will be conducted using the computer-assisted telephone interviewing (CATI) system. Help screens will be displayed with additional instructions or probes at any given point of data collection. The CATI instrument will be programmed in Blaise. Case management for the telephone interviews will utilize the Census Bureau's advanced WebCATI system. This system allows case workload to be balanced across the Census Bureau's three telephone centers and can assign cases to interviewers based on a variety of skills (e.g. language, refusal conversion expertise). As a result, cases will be handled in a more efficient and effective manner.

Optical scanning will be used to capture the digital images of the mail questionnaire after keying. The images will be stored in a database that is accessible to survey staff at their desktops. This will facilitate easier retrieval of the actual response for use during the data collection of missing critical items, data reconciliation, and editing stages.

4. EFFORTS TO IDENTIFY DUPLICATION

Duplication, in the sense of similar data collection, does not exist. No other data collection captures all components of scientists and engineers in the United States. There is no similar information available other than from this survey, conducted by the U.S. Census Bureau for NSF since the 1960s. Data from the Current Population Survey, the American Community Survey, and Decennial Census provide occupational estimates but do not collect information on degree field for higher education degrees so that these surveys cannot provide information on those with S&E degrees who are not employed in S&E occupations. Data collected in the past SESTAT surveys showed that most of those with S&E degrees (about 70%) were not employed in S&E occupations.

5. EFFORTS TO MINIMIZE BURDEN ON SMALL BUSINESS

Not applicable. The NSCG collects information from individuals only.

6. CONSEQUENCES OF LESS FREQUENT DATA COLLECTION

Because NSCG is a panel survey, conducting the survey less frequently would make it more difficult and costly to locate the persons in the sample because of the mobility of the U. S. population. The results would be a higher attrition rate and less reliable estimates. Also, government, business, industry, and universities would have less recent data to use as a basis for formulating the Nation's science and engineering policies.

Expanding the time between interviews would also lessen the accuracy of the recall of information by the respondents. This would affect the reliability of the data collected and reduce the quality of the data for all uses, including the congressionally mandated biennial reports prepared by the NSF.

Follow-up surveys every two to three years on the same sampled persons are also necessary to track changes in the science and engineering workforce as there are large movements of individuals into and out of science and engineering occupations over both business and life cycles. To make sure of the availability of current national S&E workforce data, the NSCG is conducted and coordinated with the National Survey of Recent College Graduates and the Survey of Doctorate Recipients. The degradation of any single component would jeopardize the integrity and value of the entire SESTAT system of surveys and integrated database.

7. SPECIAL CIRCUMSTANCES

Not applicable. This data collection does not require any one of the reporting requirements listed.

8. FEDERAL REGISTER ANNOUNCEMENT AND CONSULTATION OUTSIDE THE AGENCY

Federal Register Announcement

The Federal Register announcement for the NSCG appeared on March 7, 2008 (See Appendix B.) NSF received one public comment in response to the announcement as of the closeout date of May 6, 2008. The comment came from B. Sachau of Floram Park, NJ, via e-mail on March 7, 2008. Ms. Sachau objected to the information collection. Ms. Sachau had no specific suggestions for altering the data collection plans other than to discontinue them entirely. NSF responded to Ms. Sachau on March 21, 2008, describing the program, and addressing the frequency and the cost issues raised by Ms. Sachau. NSF believes that because the comment does not pertain to the collection of information on the required forms for which NSF is seeking OMB approval, NSF is proceeding with the clearance request.

Consultations Outside the Agency

The Division of Science Resources Statistics (SRS) within the NSF has responsibility for the SESTAT surveys. In the early 1990s, SRS initiated and implemented a major redesign of this system of surveys, and continued to adhere closely to the redesigned approaches in conduct of the surveys throughout the decade.

As the SESTAT survey system entered the first decade of the 21st century, SRS set a goal to further improve the efficiency and relevancy of the SESTAT system in meeting the data needs of policy makers, academic and research communities and industry. In order to accomplish this goal, SRS carefully planned and engaged in a series of formal and informal evaluations and assessments of each of the three surveys as well as the system as a whole between May 1999 and December 2002. These evaluations covered several areas: sampling frame, population coverage, sample design, survey content, data system design and data dissemination.

After the redesign efforts, SRS began a more systematic set of activities to encourage greater dissemination of the SESTAT surveys, and to encourage greater use of the data by outside researchers .

Meetings and Workshops

Both internal and external consultation took place through a series of meetings and workshops on various issues related to the SESTAT redesign and survey methodology.

For the 2003 survey round:

- SRS hosted a workshop on possible sources of alternate sampling frames for each of the three SESTAT surveys, including the NSCG, to ensure that current frames were still the most efficient and cost effective source for the populations of interest. For this workshop, representatives from the other government agencies and survey firms were invited to provide input on other potential national frames suitable for the SESTAT surveys. This workshop confirmed that there were no alternative and better frames than those used in the 1990s and still available early in the next decade.

- SRS convened an expert content panel to provide overall guidance on the review of the SESTAT questionnaire content and the relevancy of the information collected to meet policy, research and user needs. The content panel was comprised of experts knowledgeable about scientific workforce and education issues, and represented individuals from the private-for-profit industry sector, academia, and non-profit organizations. The content panel met three times (February 2000, May 2000 and June 2002); each of the meetings included invitees from other federal agencies who either collect general workforce data, or use the SESTAT data. A report was issued from each meeting. Feedback from this activity confirmed the importance of the current content of the SESTAT surveys -- the content panel members did not recommend deleting any content -- and provided guidance on new content for the upcoming decade.

- SRS commissioned the Committee on National Statistics (CNSTAT) of the National Research Council (NRC) to examine proposed sample design options for the NSCG. The CNSTAT committee held a two-day workshop on this topic, and issued a report with recommendations to NSF on the 2003 NSCG sample design. The recommendations generally were already reflected in the design plan for the surveys.[2]

For the 2006 and 2008 survey rounds:

- SRS hosted the SESTAT Data Collection Contractors Debriefing. All aspects of the SESTAT data collection were discussed to make improvements on the survey procedures.

For the 2008 survey round:

- SRS held a SESTAT Methodological Research Conference. This conference was held to share the results of the methodological research and experiments conducted in the three 2006 SESTAT component surveys.

[2] National Research Council, Committee on National Statistics. 2003. Improving the Design of the Scientists and Engineers Statistical Data System. Washington: The National Academies Press.

- SRS held the SESTAT Research Methodology Planning Meeting. This meeting was conducted to discuss and coordinate possible research experiment ideas for the 2008 SESTAT surveys.

Public Consultations on Redesign

In addition to these meetings and workshops, SRS conducted a series of consultations with the public.

For the 2003 survey round:

- SRS conducted a series of executive interviews with high-level officials in industry, government and academe to identify the important issues likely to affect the S&E workforce in the future. The interviews were conducted with eight federal officials; one state government official; nine senior staff from private, for-profit firms; six senior staff from non-profit associations or foundations; and one university administrator.

- SRS conducted telephone interviews with eleven academic and nonacademic researchers who had previously used the SESTAT data to determine how well the survey data met their needs, and to obtain feedback on other useful content areas to include on the surveys in the future.

- SRS conducted a series of five focus groups with different sets of users on all aspects of the SESTAT survey data. Two of the focus groups were held online, an innovative approach that was cost saving, as well as providing an immediate transcript. Other focus groups were conducted in a meeting setting. The focus group participants consisted of: 1) Professional association staff who routinely used the SESTAT data as a national comparison of their internal data; 2) Researchers who used the data in the past; 3) Current and past NSF staff who had been data users; 4) Current and past SESTAT contractor staff who were familiar with the data collection and areas where improvements could be made; 5) Staff from the NRC who also routinely used scientific workforce data.

- SRS commissioned five independent researchers who were familiar with the historical development of NSF's workforce data, and had extensively used the SESTAT microdata, to conduct thorough content reviews and evaluations of the survey data.

Consultations for Outreach and Dissemination

In order to maintain the currency of the SESTAT surveys and to obtain ongoing input from the public and researchers, SRS has engaged in the following activities.

For the 2008 survey round:

- SRS has convened a Human Resources Experts Panel (HREP) in order to help improve data collection on the science and engineering (S&E) labor force through review and renewal of the S&E personnel surveys and to promote use of the data for research and policy analysis purposes. HREP will accomplish its mission by: 1) Suggesting methods to publicize and promote the data; 2) Providing advice on efforts to improve the timeliness and accuracy of S&E labor force data; 3) Providing a mechanism for obtaining ongoing input from both researchers and policy analysts interested in S&E personnel

data; 4) Providing perspectives on the data needs of decision makers; 5) Identifying issues and trends that are important for maintaining the relevance of the data; 6) Identifying ways in which S&E personnel data could be more useful and relevant for analyses; and 7) Proposing ways to enhance the content of the SRS human resources surveys. The panel is made up of 15 members who represent the sciences, academia, business/industry, government, researchers and policy makers. The panel is scheduled to meet twice a year for 3 years. Two meetings have been held since the panel convened in 2007.

- In addition to researchers and the public who use the public-use SESTAT, SDR, NSRCG or NSCG files, there are also individuals who use the restricted-use files under a license. SRS has funded two workshops over the past 18 months where a selection of current and potential future licensees met at NSF to present their research findings to NSF as well as to the broader research community.

- The SESTAT surveys contain a wealth of information on highly-trained individuals in the U.S. labor force. Over the past several years, there has been a great deal of interest in leveraging the survey data that are collected with other information on productivity by some of the same individuals (for example, patenting records or publishing records). In order to pursue the feasibility of this approach, SRS funded a workshop at NSF that brought in experts on database matching.

- As part of broader SRS activity on innovation, SRS participated in the 2007 Symposium on Entrepreneurship and Innovation Data. The purpose of the workshop was to present researchers with information on datasets that could advance research in this area.

9. PAYMENT OR GIFTS TO RESPONDENTS

Because the NSCG interviews the same individuals over time, NSF is concerned that offering respondents monetary incentives in one cycle will have an adverse effect on their survey responses in subsequent survey cycles. To better understand the effect of incentive conditioning, a $10 prepaid monetary incentive experiment is planned for the sample of panel members who received a monetary incentive in the previous survey cycle. In addition to an incentive conditioning experiment, NSF plans to offer a $20 prepaid monetary incentive to a sample of nonrespondents near the end of the data collection period to minimize potential nonresponse bias. See sections B.3 and B.4 for details on the incentives.

10. ASSURANCE OF CONFIDENTIALITY

NSF and the Census Bureau are committed to protecting the anonymity of all survey respondents. The NSCG data will be collected in conformance with the Privacy Act of 1974, NSF's authorizing legislation and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002.

As explained in Section B.1, there are two components of the 2008 NSCG sample design. The first one is the 2003 NSCG respondents from the 2000 decennial census, while the other one is the "NSRCG panel" respondents subsampled from the 2001, 2003, and 2006 NSRCGs.

The statement on the questionnaire cover will cite the data collection authority as the NSF Act and confidentiality assurances under the CIPSEA. The questionnaire cover statement will also inform the respondents that the data will be used for statistical purposes only, and the voluntary nature of their response. Two different cover letters will be used for the 2003 NSCG respondent sample and for the NSCG "NSRCG panel" sample. For the 2003 NSCG respondents, the cover letter will include additional statements about the Census Bureau's Title 13 as the data collection authority and assurances of confidentiality (see Appendix E). The Census Bureau will include the same appropriate notices of confidentiality and the voluntary basis of the survey in the introduction to respondents contacted during the CATI phase of the data collection.

NSF and the Census Bureau will operate within the guidelines established by the Privacy Act to protect respondents' privacy and the confidentiality of the data collected. The Privacy Act states that "microdata files prepared for purposes of research and analysis are purged of personal identifiers and are subject to procedural safeguards to assure anonymity."

The Census Bureau has demonstrated experience in handling sensitive data. Routine procedures will be in place to ensure data confidentiality, including the use of passwords and encrypted identifiers to prevent direct or indirect disclosures of information. Furthermore, the Census Bureau's management system is in full compliance with the government's ADP systems requirements.

11. JUSTIFICATION FOR SENSITIVE QUESTIONS

No questions of a sensitive nature are asked in this data collection.

12. ESTIMATE OF RESPONDENT BURDEN

The NSF estimates that it will contact approximately 68,000 sample persons by mail or computer-assisted interviewing. Based on experience administering the NSCG interviews, the questionnaire takes an average of 25 minutes to complete. With two modes of data collection, an overall response rate of about 90 percent is estimated. Based on an estimate of approximately 61,200 completed cases, the total burden hours for the 2008 NSCG are 25,500 for the main data collection. Additionally, about 100 burden hours are estimated for future testing of methods to reduce burden and improve utility for the 2010 survey. The total cost to respondents for the 25,600 burden hours is estimated to be $775,424. This estimate is based on an estimated median annual salary of $63,000 per NSCG respondent. Assuming a 40-hour workweek and a 52-week salary, this annual salary translates to an hourly salary of $30.29. Salary estimates were obtained using data from the 2006 NSCG.

13. COST BURDEN TO RESPONDENTS

Not applicable. This survey does not require respondents to purchase equipment, software or contract out services.

14. COST BURDEN TO FEDERAL GOVERNMENT

The total estimated cost to the Government for the 2008 NSCG is approximately $9.3 million, which includes survey data collection costs, NSF staff costs to provide oversight and coordination with the other two SESTAT surveys, and costs associated with the integration of NSCG data into the SESTAT data system. The cost estimate for data collection is $8,500,000, which is based on sample size; length of questionnaire; administration; overhead; sample design; mailing; printing; sample person locating, telephone interviewing; incentive payments, critical items data retrieval, data keying and editing; data quality control; imputation for missing item responses; weighting and estimating sampling error; file preparation and delivery; and preparation of documentation and final reports. The NSF staff costs are estimated at $417,000 (based on $111,104 annual salary of 1.5 FTE for 2.5 years of the 2008 NSCG survey cycle). The SESTAT integration costs is estimated at $400,000 for the 2008 NSCG survey cycle.

15. REASON FOR CHANGE IN BURDEN

There will not be much change in burden from the 2006 survey because the sample size for the 2008 survey will remain the same at 68,000. The 2008 survey will attempt to interview the same individuals who responded to the 2006 NSCG, and a subsample of recent graduates who responded to the 2006 NSRCG. The only change in the burden hours from the 2006 NSCG is inclusion of additional burden hours required for future tests to improve survey procedures and utility.

16. SCHEDULE FOR INFORMATION COLLECTION AND PUBLICATION

NSF does not plan to use any complex analytical techniques in NSF publications using this data. Normally cross tabulations of the data are presented in NSF reports and other data releases.

The time schedule for 2008 data collection and publication is currently estimated as follows:

Data Collection	October 2008 - March 2009
Coding and Data Editing	December 2008 - August 2009
Final Edited/Weighted/Imputed Data File	November 2009
SESTAT Info Brief	Spring 2010
SESTAT Detailed Statistical Tables	Summer 2010
SESTAT Integrated Public Use Data File	Summer/Fall 2010

17. DISPLAY OF OMB EXPIRATION DATE

The OMB Expiration Date will be displayed on the 2008 NSCG questionnaire.

18. EXCEPTION TO THE CERTIFICATION STATEMENT

Not Applicable.

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

The sampling frame for the 2008 NSCG will include approximately 75,000 cases that originated from the 2003 NSCG, the 2001 NSRCG, the 2003 NSRCG, and the 2006 NSRCG. Individually, these four surveys collected information on degrees earned prior to April 1, 2000, between April 1, 2000 – June 30, 2000, between July 1, 2000 – June 30, 2002, and between July 1, 2002– June 30, 2005, respectively. Combined, these surveys collected information on degrees earned prior to June 30, 2005.

The 2003 NSCG cases originated from the 2000 Decennial Census long form sample. The 2003 NSCG sample design can be characterized as a stratified design with probability-proportion-to-size (PPS) systematic selection using the Long Form sampling weight. The 2001, 2003, and 2006 NSRCG cases originated from a two-phase design that sampled postsecondary institutions and recent cohorts of graduates within the sampled institutions.

To be included in the 2008 NSCG frame, the respondent had to have been living in U.S., have at least one bachelor's degree in an science and engineering (S&E) field, or have a least a bachelor's degree in a non-S&E field but work in an S&E occupation as of the reference week of the originating survey, and be under age 76 as of the reference week of the 2008 survey. The sample universe will cover the United States, Puerto Rico, and the U.S. territories. Approximately 68,000 persons will be selected for the 2008 NSCG sample.

The 2008 NSCG sample design will be similar to previous NSCG survey cycles. In the 2008 NSCG sample design, 2006 NSCG respondent cases that originated in the 2003 NSCG, 2001 NSRCG , or 2003 NSRCG will be sampled with certainty. Respondent cases that originated in the 2006 NSRCG will be sampled using the PPS sample selection methodology.

The targeted overall weighted response rate on the 2008 NSCG is 90 percent. The initial survey year weighted response rates for the 2003 NSCG, the 2001 NSRCG, 2003 NSRCG, and 2006 NSRCG were 73 percent, 79 percent, 68 percent, and 66 percent, respectively (these are the four sampling frame source surveys for the 2008 NSCG). Only the respondents in the previous survey cycle were followed in the 2006 NSCG and together achieved an 87 percent response rate. The plan for maximizing the response rate is presented in Section 3.

2. STATISTICAL PROCEDURES

The 2008 NSCG sample will be stratified by frame source (2003 NSCG, 2001 NSRCG, 2003 NSRCG, and 2006 NSRCG), demographic group, highest degree type, highest degree field, occupation, and sex. The demographic group is a composite variable recording disability status, citizenship, and race/ethnicity. As noted above, 2006 NSCG respondents from the 2003 NSCG, 2001 NSRCG, and 2003 NSRCG will be sampled with certainty. Approximately 50% of the respondents to the 2006 NSRCG will be included in the 2008 NSCG sample. The sample allocation of the 2006 NSRCG portion is designed to bring the sampling weights of these cases in line with the weights of cases from the 2001 NSRCG and the 2003 NSRCG.

The 2006 NSRCG portion of the 2008 NSCG will be selected using sampling strata based on a multi-way cross of the stratification variables. (See Appendix C for the 2008 NSCG sampling strata.) The 2008 NSCG sample size and sample design ensure NSF will maintain the ability to produce the small demographic/degree field estimates that are needed for the Congressionally mandated report on *Women, Minorities and Persons with Disabilities in Science and Engineering* (See 42. U.S.C., 1885d).

Estimates from the 2008 NSCG will be based on standard weighting procedures. As was the case with sample selection, the weighting adjustments will occur separately for cases from each originating survey. Each case will have a base weight defined as the probability of selection into the 2008 NSCG sample. This base weight will reflect the differential sampling across strata. Because the 2003 NSCG, 2001 NSRCG, and 2003 NSRCG respondents to the 2006 NSCG will be selected for sample with certainty, the base weight will be equal to the final weight from the previous survey cycle. Base weights will be adjusted for nonresponses. After weights are adjusted for nonresponses, weights will then be raked to ensure that the original sampling stratum totals agree with the population totals.

Replicate Weights. A set of replicate weights based on the successive difference and jackknife replication methods will also be constructed. The entire weighting process applied to the full sample will be applied separately to each of the replicates to produce a set of replicate weights for each record.

Standard Errors. The successive difference and jackknife replication methods will be used to estimate the standard errors of the 2008 NSCG estimates as in the past. The variance of a survey estimate based on any probability sample may be estimated by the method of replication. This method requires that the sample selection, the collection of data, and the estimation procedures be independently carried through (replicated) several times. The dispersion of the resulting estimates then can be used to measure the variance of the full sample.

3. METHODS TO MAXIMIZE RESPONSE

Maximizing Response Rates

In order to maximize the overall survey response rate, NSF and the Census Bureau will implement procedures such as conducting extensive locating efforts and follow-up telephone interviews for nonrespondents to the mail questionnaire. The contact information obtained from the 2006 NSCG and 2006 NSRCG for the sample members and for the people who are likely to know the whereabouts of the sample members will be used to locate the sample members in 2008.

The Census Bureau will use a combination of locating and follow-up methods similar to the procedures used for the 2006 NSCG to maximize the survey response rate. The Census Bureau will utilize all of the available locating tools and resources to make the first contact with the sample person. The Census Bureau will use the U.S. Postal Service (USPS)'s automated National Change of Address (NCOA) database to update addresses for the sample. The NCOA

incorporates all change of name/address orders submitted to the USPS nationwide, which is updated at least biweekly.

Prior to mailing the questionnaires, the Census Bureau's National Processing Center will engage in locating efforts to find good addresses for problem cases. The questionnaire mailings will utilize the "Return Service Requested" option to ensure that the postal service will provide a forwarding address for any undeliverable mail. The locating efforts will include using such sources as educational institutions and alumni associations, Directory Assistance for published telephone numbers, Phone Disc for unpublished numbers, FastData for address searches, and local administrative record searches such as researching motor vehicle department records. Private data vendors also maintain up to 36-month historical records of previous address changes. The Census Bureau will utilize these data vendors to ensure that the contact information is up-to-date.

Incentive plan in 2008

To increase the response and minimize potential bias, a $20 prepaid monetary incentive will be offered to a sample of nonrespondents near the end of the data collection. The NSF conducted several incentive experiments in the 2006 NSCG which found that an incentive offered to a sample of late respondents, including refusals at the end of the field period significantly improved their response rate compared to those who were not offered an incentive[3].

The refusals in the survey are likely to remain as nonrespondents without an offer of incentive. Incentives are effective in increasing the survey response rate, which in turn help to minimize possible nonresponse bias in the final survey estimates.

We propose to follow the same procedures used in the 2006 survey by offering a $20 prepaid incentive to the late respondents as a gaining-cooperation strategy. The sample selected to receive the incentive will be determined after implementation of the incentive conditioning experiment and the standard survey data collection protocols.

The overall survey response rates and the number of respondents at three months prior to the end of the field period will be analyzed by sampling cell. The overall strategy would be to give all nonrespondents a probability of receiving an incentive. A greater probability of selection for receiving an incentive will be given to cases in those cells where the number of completed cases is low, in order to improve the survey estimates.

We will examine each sampling cell, and apply the following allocation formula to determine the number of cases in each sampling cell that will receive the incentive.

[3] See Kinnaman, Deborah, "Results of 2006 NSCG and NSRCG Postpaid Incentives Experiments", U.S. Bureau of the Census Memo to the National Science Foundation, December 2007.

$$n_h = n * \frac{\frac{p_h^a}{r_h^b}}{\sum_{h=1}^{H} \frac{p_h^a}{r_h^b}}, \text{ here } a \geq 0, b \geq 0 .$$

where:

n_h = is the number of incentives allocated to sampling stratum h

n = number of incentives available (based on budget) for this survey

p_h = nonresponse rate for stratum h before the incentives are issued

r_h = number of respondents before the incentives are issued

H = total number of sampling strata

a, b = parameters used to determine whether p_h or r_h should be given more weight.

NSF will ensure that all nonrespondents have a 0.25 up to 1.0 probability of receiving the incentive. The final determination of the incentive distribution will be made in consultation with the SRS Chief Statistician. The incentive will be offered via U.S. Priority Mail with a paper questionnaire and during the telephone contacts of sample members in the incentive group. The date of this incentive offer would likely be in early January 2009, which is about two months prior to planned survey field closeout.

Dealing with Issues of Nonresponse Bias

Traditionally, the response rate on the first postcensual survey is lower than the subsequent follow-up surveys due to various reasons. The 1993 NSCG weighted response rate was 80 percent but subsequent surveys had response rates far above 90 percent. The NSCG weighted response rate was 73 percent in 2003 and 87 percent in 2006.

NSF was concerned with the lower than expected NSCG response rate in 2003, and took several measures to evaluate and address potential nonresponse bias in the 2003 data. NSF asked the Census Bureau to conduct a detailed nonresponse bias analysis. NSF also contracted an independent analysis of the 2003 NSCG data, which identified significant differential response rates by age of sample members where younger age groups were much more likely to be nonrespondents to the survey than older age groups.

The Census Bureau issued nonresponse reports on unit and item nonresponse rates in the 2003 and 2006 NSCG data by various respondent and nonrespondent characteristics and data collection stages. Results from the nonresponse research and analysis were used extensively in the nonresponse weighting adjustments to reduce the nonresponse bias in the 2003 and 2006 NSCG data. Careful selection of factors for constructing the weighting classes were done to reduce possible nonresponse bias. Weights were also adjusted to control distributions for some variables to known totals from the sample frame.

In 2008, further assessment will be made of the extent of remaining bias by comparing weighted estimates for the survey sample that can be observed in the sample frame (e.g. degree field,

degree level, and gender) to estimates for the population that the weighted sample is intended to represent.

4. TESTING OF PROCEDURES

Because data from all three SESTAT surveys are combined into a unified data system, the surveys must be closely coordinated to provide comparable data from each survey. Most questionnaire items in the three surveys are the same.

Although there will be no new questions in the 2008 NSCG questionnaire, all content items in the SESTAT questionnaires have undergone an extensive review and testing before they were included in the final version. The changes made in the questionnaires are a result of a variety of activities that included extensive review of the entire content in each of the SESTAT survey questionnaires and additional research on specific items to provide more information before a final decision was made on placement and wording of the item in the questionnaires. Content evaluation and testing activities for the 2003 and 2006 surveys included:

- External and internal consultation with questionnaire design experts on questionnaire layout and formatting to improve user-friendliness and minimize respondent reporting errors;
- External consultation on improving the messages in the survey contact materials; and
- A two-stage pretest of the survey questionnaires consisting of mail and telephone.

All of these activities contributed to the development of the questions in the NSCG questionnaire.

Survey Questionnaire Review and Research

The SESTAT survey questionnaire items are divided into two types of questions: core and module. Core questions are defined as those considered to be the base for all three SESTAT surveys. These items are essential for sampling, respondent verification, basic labor force information, and/or robust analyses of the science and engineering workforce in the SESTAT integrated data system. They are asked of all respondents each time they are surveyed, as appropriate, to establish the baseline data and to update the respondents' labor force status and changes in employment and other demographic characteristics. Module items are defined as special topics that are asked less frequently on a rotational basis of the entire target population or some subset thereof. Module items tend to provide the data needed to satisfy specific policy, research or data user needs.

After identifying the core and module items that would be included in the SESTAT surveys, SRS reviewed and identified content items needing improvement, and engaged in research to craft new questions. SRS conducted separate studies on six core items, and one study on a module for the 2003 survey questionnaires. The core item research covered the following topics on the SESTAT questionnaires: employer's main business, academic positions, academic institutions, work activities, marital status, and degrees earned abroad.

The core item research resulted in some wording changes to those questions on the SESTAT questionnaires, and a revision of how the occupation code frame is presented. The 2008 NSCG questionnaire will not include new questions not previously fielded before.

For 2008, the NSCG questionnaire content will be revised from 2006 as follows:

- Survey reference date changed from April 1, 2006 to October 1, 2008.
- Removed a 2006 module on collaborative activities (it has not yet been decided if this will be rotated back in at a future time).
- Rotated in a module on second job (status, job description, job category, relatedness of second job to highest degree), which was asked in 1993-1999.
- Rotated in a module on respondent's and spouse's areas of technical expertise, which was asked in 1993-2003.

A complete list of questions proposed to be added, dropped, or modified in the 2008 NSCG questionnaire is included in Appendix D.

The 2008 NSCG questionnaire retains all content changes that were tested and implemented for the 2006 SESTAT questionnaires. In 2005, SRS conducted an extensive pretest under a generic clearance (OMB No. 3145-0174) that consisted of two phases: (1) two rounds of in-depth cognitive interviews, and (2) a small-scale field test of the mail questionnaires.

Pretest Phase I – Cognitive interviews

Mathematica Policy Research, Inc. (MPR) and the U.S. Census Bureau (Survey Research Division) were contracted to conduct in-depth cognitive interviews on the 2006 NSCG and the other two SESTAT survey questionnaires. Cognitive interviews were conducted in two waves, with the waves being scheduled during the same time period at MPR and the Census Bureau. MPR tested the full-length questionnaires for the three surveys, while the Census Bureau was asked to focus on the employment section of the NSCG. In addition to the questionnaires, the cognitive interviews were also used to test improvements to the cover letters for the 2006 survey administration.

The first round of cognitive interviews was conducted between February 2 and February 25, 2005. During this period MPR and Census Bureau each interviewed 30 respondents. The second round of cognitive interviews was conducted between March 25 and May 2, 2005. MPR interviewed 40 respondents (28 in-person and 12 via telephone) and the Census Bureau interviewed 30 respondents. Based on the results of the cognitive interviews, MPR and NSF worked together to develop a series of experiments to test in the mail portion of the pretest.

Pretest Phase II – Mail Field Test

The field test consisted of two mailings of NSCG and the other two SESTAT surveys with a reminder postcard in between; no further nonresponse follow-up was conducted due to time constraints. The NSCG mail pretest included a sample of 1,500 selected from a commercial list of 5,000 names of bachelor's degree holders with address, sex, age, and occupation information, and between the ages of 21 and 75. To mimic the proportion of science and engineering cases

from the 1995 NSCG, MPR selected 15 percent of the cases from computer occupations, 20 percent from engineering occupations, and 65 percent from other occupations for a total of 1,500 sample members. Each sample member was randomly assigned to one of four control or experimental groups.

Pretest questionnaires were mailed on June 24, 2005 using first class mail. Although mailing a reminder was not part of the original pretest plan, a postcard reminder was sent to all non-respondents because of the low response (12 percent) to the first mailing. The postcard was mailed on July 20, 2005, and provided an additional boost of about 2 percentage points to the response rate for a 14 percent cumulative overall response rate from all three SESTAT surveys to the first mailing. A second mailing was sent on August 3, 2005 with a cover letter urging participation with a "respond by" date in a Priority Mail envelope. Mail returns were accepted until August 26, 2005. Final response rate to the NSCG mail pretest was about 25%. Final response rate for respondents from all three surveys was 27 percent.

The primary goal of the field pretest was to test the various recommended questionnaire changes from the cognitive interviews. Specific test conditions were incorporated to obtain research data that might further improve the questionnaires. These are described below:

1) Testing the placement of the sample person's name and address label on the questionnaire (front versus back cover).
2) Testing the Field of Study and Job Category Code Lists in a new format.
3) Testing a different approach to "anchoring" the reference date in the employment questions.
4) Testing a new wording and format of the principal employer type question.

In addition, the experimental versions of the questionnaires had small wording and formatting changes for some questions of interest such as work activity categories, employer name and location, supervising, etc. The control versions of the questionnaire retained the same wording for most questions of interest and Field of Study/Job Category Code Lists used in 2003. Testing the label placement by the presence versus absence of the content changes created a two-by-two design, shown in table below.

Mail Pretest Design

		Content, Anchor, and Code List	
		Old Content (Control)	New Content (Experimental)
Address Label	Back	Questionnaire Version 1	Questionnaire Version 3
	Front	Questionnaire Version 2	Questionnaire Version 4

The mail pretest also included testing of a new 2006 module on the method and means of collaboration; using "Yes/No" response options in a few remaining questions with the "Mark All That Apply" response options used in 2003; moving the part-time employment questions to a different section and revising the work-related training reasons to fine tune the measurement of the concepts for these two items.

Based on the mail pretest results, decisions were made to keep the sample person's name and address labels on the front cover of the questionnaire; use the revised wording and format of the employer sector question; use the new Field of Study/Job Category Code Lists; no longer use the 'Mark All That Apply' response option; not use the reference week "anchoring" question but use consistent question wording in all references to the principal job.

Survey Contact Materials

The cover letters for the 2008 NSCG questionnaire will be developed based on the results from the 2003 NSCG Cover Letter research which tested the impact of different cover letters. This research showed a marginal response increase with the new "altruistic" cover letter overall and "authoritative" cover letter was found to be effective among respondents in some fields. These two types of cover letter will be used again as the main letters to the sample members in 2008 (Appendix E).

Questionnaire Layout

SRS has previously engaged the services of Dr. Don Dillman to further improve the visual presentation of the 2003 and 2006 SESTAT questionnaires. An SRS staff member with expertise in visual design theory was also involved in this process. The suggested revisions to the questionnaires included the standardization and consistent use of formatting, placement of instructions, and placement of privacy act notices. Also revised were the items that include a format that requires the respondent to review a long list of items before reporting a response to make the selection process easier for the respondents.

2006 Survey Methodology Tests

<u>Postpaid Incentive Experiment</u>

In 2006, the Bureau of the Census conducted a postpaid incentive experiment in the NSCG. This experiment was designed to increase the response rate of the late respondents who were either classified as refusals (both soft and hard), targeted nonrespondents (NSCG "RCG panel" sample cases had significantly lower response rate than the 2003 NSCG decennial cases), and elusive nonrespondents (contact information confirmed to be correct but cannot reach the sample person) by offering a postpaid monetary incentive in the form of an unactivated $20 VISA gift card. Once the interview was completed, the respondents were told that the gift card would be activated within two business days. This unactivated card was included in the final questionnaire mailing and also offered during the CATI calls to the incentive treatment group of respondents. There was also a control group that did not receive an incentive.

The experiment found that the incentive increased the response rate about 17% for previous NSCG refusal cases, 14% for targeted nonrespondents, and 11% for elusive nonrespondents. The differences in the response rates between the incentive and control groups were statistically significant.

Reminder Experiment

This experiment tested four different means of reminding mail recipients to return their questionnaires. The purpose of this experiment was to determine the best reminder method for the 2006 NSCG. The methods tested were the traditional Dillman postcard reminder method, a letter reminder, an automated telephone reminder, and an email reminder. The experiment showed that no one reminder method was more effective than any other at increasing response rates. The 2008 NSCG will use postcard, email and telephone reminders through the data collection phase because each reminder had an immediate effect in boosting the survey responses when administered.

Due Date Contact Experiment

This experiment tested whether a request to "Please complete and return within two weeks" (due date) notice encourages a faster survey response than "Return as soon as possible" statement typically used in the survey contact materials. An increase in early response by mail would decrease the follow up workload and thus survey cost. Four groups consisted of due date notice only on the envelope; due date notice on the cover letter only; due date notice on both the envelope and cover letter; and the control group that had "return as soon as possible" notice. The experiment showed that the group with the due date notice on both the envelope and cover letter had the highest early response rate of all groups. The NSCG will include the due date notice on both the envelope and cover letters in 2008.

Survey Methodology Tests to be Undertaken

As described in Section A, to better understand the effect of incentive conditioning on the survey panel, a monetary incentive experiment is proposed for the sample members who received a monetary incentive in the 2006 survey round. The 2006 NSRCG panel sample in the 2008 NSCG, who received $5 or $10 prepaid incentives with the first questionnaire mailing in 2006, will be split into treatment and control groups where only the treatment group will again receive a $10 prepaid incentive with the first questionnaire mailing. The incentive experiment is designed to determine if the previous incentive receipt has any negative effect on the subsequent survey participation when no incentive is offered. Details on the incentive conditioning experiment plan are in Appendix G

NSF plans to conduct additional methodological tests in the current and future rounds of the survey to reduce burden and increase utility of the survey under the burden hours in this survey clearance for the next survey cycle. Proposals for these additional tests are still under consideration. These will be submitted for OMB approval prior to implementation.

5. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION

Chief consultant on statistical aspects of data collection is John M. Finamore (301) 763-5992, Demographic Statistical Methods Division, Census Bureau. The Demographic Statistical Methods Division will manage all sample selection operations at the Census Bureau. At NSF the contacts for statistical aspects of data collection are Stephen Cohen, SRS Chief Statistician (703) 292-7769, and Kelly Kang, NSCG Project Manager (703) 292-7796.

APPENDIX G

2008 NSCG Incentive Conditioning Experiment

Proposal for the 2008 NSCG Incentive Conditioning Experiment

Background

Providing incentives has been shown to increase survey response rates and is used often in survey research. However there could be unintended consequences to the use of incentives, particularly when the same respondents are interviewed subsequent times. Respondents could develop an expectation of payment for participation in the survey (Singer, VanHoewyk and Maher, 1998).

Cotterchio and Krieger (1998) found that cash incentives improved response rates to their main study. In a follow-up study, however, response rates for those who received an incentive during the first study were lower than for those who did not receive an incentive during the first study. They purport that the study subjects may have expected to receive an incentive for participation and when no incentive was provided for the second part of the study, were less likely to participate.

Singer, VanHoewyk and Maher (1998) found that while respondents who had previously been given an incentive were more likely to agree with the statement that "people should be paid for doing surveys like this," they were more likely to participate in a later round of the survey even without an incentive payment.

As stated by Singer, Groves and Corning (1999), if a respondent has an expectation of payment for survey participation, a violation of that expectation by the survey organization could lead to reluctant participation or refusal to participate. In a 1996 study that looked at respondents' opinions about differential incentives, they found no significant difference in willingness to participate in future survey rounds between those who thought differential incentives were unfair and other respondents.

During previous rounds of the National Survey of College Graduates (NSCG) and National Survey of Recent College Graduates (NSRCG), some respondents have been provided monetary incentives for their participation.

Since the NSCG has never included an experiment to measure the effect of incentive conditioning and since outside research conducted has not reached a consistent conclusion on the conditioning effect of incentives, we would like to evaluate this effect on the 2008 NSCG sample. In general, we hope to learn whether participation in future rounds of the survey is impacted by previous experience receiving an incentive. In particular, we would like to answer the following question:

- If a person receives an incentive in the initial year to complete a survey, does not receiving an incentive in a subsequent year negatively affect their response propensity?

To gain a better understanding of the response behaviors related to incentives, additional questions we would like to answer through this research include the following:

- If a person receives an incentive in the initial survey year to complete a survey, will they respond more favorably to an incentive in a subsequent year than someone who had not received an incentive in the initial year?
- Are incentives more effective the initial time they are used than if they are given in subsequent years?

We will attempt to answer these questions through the comparisons listed in the methods section.

Methods

The 2008 NSCG Incentive Conditioning Study will have four treatment groups. In the 2006 National Survey of Recent College Graduates (NSRCG) a sample of cases in traditionally low responding majors were randomly assigned to receive a $5 or $10 prepaid incentive.

Cases will be selected from 2006 NSRCG cases that were eligible for the 2006 prepaid incentive study that either did or did not receive a prepaid incentive. Cases that were eligible for the 2006 NSRCG **postpaid** incentive study will **not** be eligible for the 2008 NSCG incentive conditioning study. Additionally, cases that do not have a valid address as of the 2008 NSCG data collection planning period will be ineligible. The 2006 NSCG incentives and no incentives cases will each be split into two groups for 2008: incentive and no incentive.

Breaking the study cases out this way will create the following 2 by 2 factorial design with four treatment groups.

Treatment		2008 Incentive Status (factor B)	
		Yes	No
2006 Incentive Status (factor A)	Yes	Group 1: (Yes, Yes)	Group 2: (Yes, No)
	No	Group 3: (No, Yes)	Group 4:(No, No)

- Group 1 – Received **an incentive** in 2006, and will receive **an incentive** in 2008 (Yes, Yes)
- Group 2 – Received **an incentive** in 2006, but will receive **no incentive** in 2008 (Yes, No)
- Group 3 – Received **no incentive** in 2006, but will receive **an incentive** in 2008 (No, Yes)
- Group 4 – Received **no incentive** in 2006, and will receive **no incentive** in 2008 (No, No)

With this design, a two way ANOVA will be used to analyze the data. A graphic for the resulting response rates for the four treatment groups will be made to help in visualizing the effects (see graphic 1 below as an example). This will allow us to study the main effects, make group comparisons, and study the interaction. A few examples of the comparisons of interest are:

Main effects:

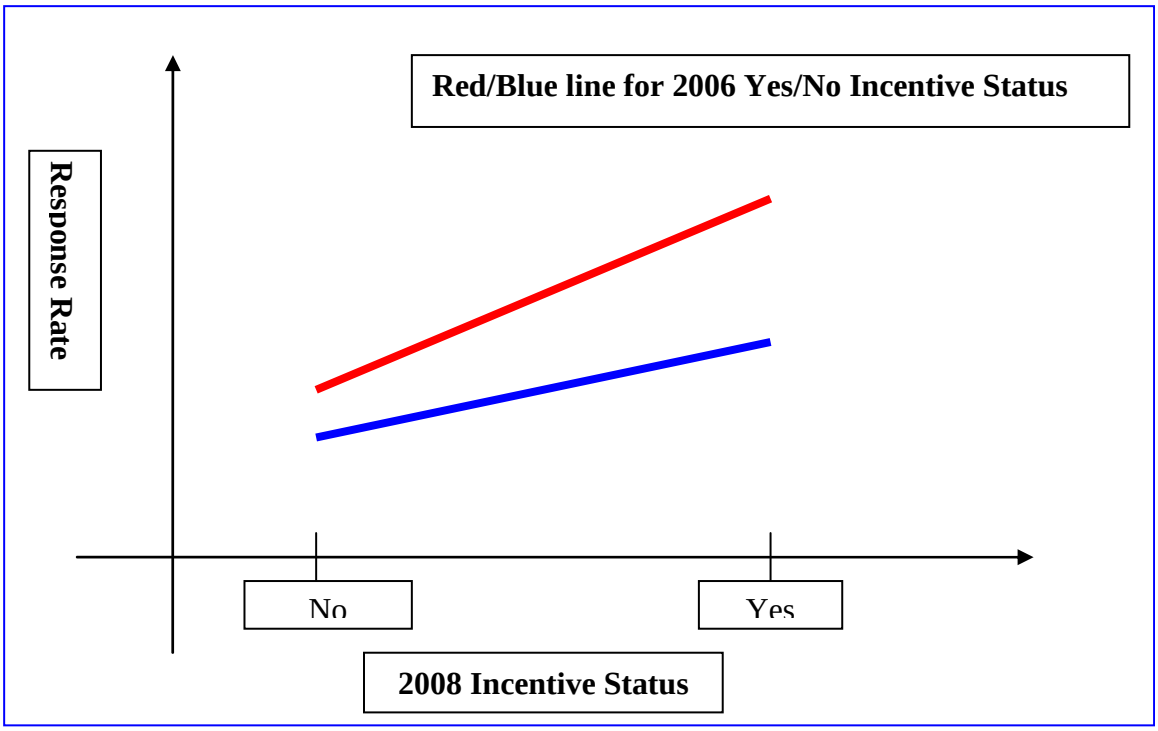
- Group 1 (Yes, Yes) + Group 3 (No, Yes) vs Group 2 (Yes, No) + Group 4 (No, No)
- Group 1 (Yes, Yes) + Group 2 (Yes, No) vs Group 3 (No, Yes) + Group 4 (No, No)

Group Comparisons:

- Group 1 (Yes, Yes) vs. Group 2 (Yes, No) – If a person receives an incentive in the initial survey year to complete a survey, how will receiving or not receiving an incentive in a subsequent year affect their response tendency?
- Group 2 (Yes, No) vs. Group 4 (No, No) – How will receiving or not receiving an incentive in the initial survey year affect a person's response tendency if they do not receive an incentive in a subsequent year?

Interactions:

- [Group 1 (Yes, Yes) - Group 2 (Yes, No)] vs. [Group 3 (No, Yes) - Group 4 (No, No)] – What is the relative effect of incentives when they are offered a second time versus the first time they are offered?



Cases assigned to receive an incentive in 2008 will have a prepaid $10 incentive gift card included in the initial questionnaire mailing. We will provide $10 because we do not want to provide a lower incentive to those who received $10 in the 2006 NSRCG. Cases assigned not to receive an incentive will receive the same initial questionnaire mailing as the main group of NSCG cases that were not eligible for this study. All of the cases in the study will start in the mail data collection mode and will follow the same mailing procedures as the main NSCG cases. At the start of the Computer Assisted Telephone Interview (CATI) nonresponse follow-up, these cases will be moved into the CATI data collection phase along with the other NSCG cases.

Limitations

One study limitation is that we are restricted to those who have received a prepaid incentive in a previous survey round. Due to the prepaid incentive study design from the 2006 NSRCG, we are limited to those NSRCG respondents who are in traditionally low responding majors. As a result, the incentive conditioning study findings may not be generalizable to the entire NSCG target population.

Another limitation is that during the 2006 NSRCG, prepaid incentives were provided in either $5 or $10 amounts. Due to sample size limitations, we will put both of these groups into the same group of people who received an incentive in 2006. Respondents could possibly receive a greater incentive than they did in 2006, which could cause differences in response rates.

Sample Size Calculations

The sample size for each treatment group was calculated based on the minimum number of cases needed in each treatment to detect a 10% difference in response rate. The sample size formula was as follows:

$$n \geq (Z_{\alpha^*/2} + Z_{\beta})^2 \frac{p_1(1-p_1) + p_2(1-p_2)}{\delta^2} \qquad (1)$$

Where:

n = sample size for a single treatment group
α^* = alpha level adjusted for multiple comparisons
$Z_{\alpha^*/2}$ = critical value for set alpha level assuming a two-sided test
Z_{β} = critical value for set beta level
p_1 = proportion for group 1
p_2 = proportion for group 2
δ = minimum detectible difference

Since there are seven comparisons included in this study (one main comparison and six supplemental comparisons), the alpha level used in the sample size calculations was adjusted for multiple comparisons. Using the Bonferroni adjustment, the Census Bureau standard alpha level of 0.10 was decreased to account for the seven comparisons. Therefore, the alpha level in the sample size calculations was 0.10/7 = 0.014.

The beta level was included in the formula to inflate the sample size in order to decrease the probability of committing a type II error. Committing a type II error, claiming there was no difference in response rates across mode groups when a difference was present, would be detrimental to the purpose of this study. With this in mind, the beta level was set to 0.90.

The estimated response rate for the groups was set to 0.50 for the sample size calculations. Setting the estimated response rate at this value was a conservative approach that maximized the sample size. The final term needed for the sample size calculation was the desired minimum detectible difference. We would like to be able to detect a 10% difference in response rate of any comparison.

Using the values discussed above, the sample size calculation was as follows:

$$n \geq (2.452 + 1.282)^2 \frac{(0.50(1-0.50) + 0.50(1-0.50))}{(0.1)^2} = 697.13$$

This sample size tells us that in order to consider a 10% difference in response rate in any comparison as statistically significant, we need at least 698 sample cases in each of the four treatment groups. The following tables show that there will be an adequate number of cases available to achieve our sample sizes.

The eligible cases for the 2008 NSCG incentive conditioning study are cases that were eligible for the 2006 NSRCG prepaid incentive but not eligible for the 2006 NSRCG postpaid incentive. Cases

eligible for the postpaid incentive were refusals or late responders. Excluding these cases will help simplify our design and analysis since we will be dealing with cases that are not late responders and have not received multiple incentives. Additionally, cases with bad addresses will be ineligible for the study. Table 1 provides a breakdown of the cases considered for the 2008 NSCG Incentive Conditioning Study. The highlighted rows in the table identify the cases that are eligible for the study.

Table 1. Number of Eligible Cases for 2008 NSCG Incentive Conditioning Study

2006 Prepaid Incentive?	Eligible for 2008 Incentive Conditioning Study	Frequency
Yes	Ineligible – Part of 2006 NR Incentive	268
	Ineligible – Part of 2006 Refusal Incentive	53
	Ineligible – Bad Address	210
	Eligible for 2008 Incentive Conditioning Study	1,891
No	Ineligible – Part of 2006 NR Incentive	540
	Ineligible – Part of 2006 Refusal Incentive	115
	Ineligible – Bad Address	286
	Eligible for 2008 Incentive Conditioning Study	2,732
Total		**6,095**

For the 2008 NSCG incentive conditioning study there are four treatment groups.

- We will split the group that received a 2006 NSRCG prepaid incentive (1,891 cases) into two groups: one that will receive an incentive and one that will not receive an incentive as part of the 2008 NSCG.
- We will split the group that **did not** receive a 2006 NSRCG prepaid incentive (2,732 cases) into two groups: one that will receive an incentive and one that will not receive an incentive as part of the 2008 NSCG.

As derived using formula (1), the sample size for each of the four treatment groups will be at least 698 cases. Using the figures from Table 1, there is an adequate number of cases eligible for the study to ensure at least 698 cases in each treatment group. The table below documents the eligible number of cases and the number needed for this incentive conditioning study.

Table 2. 2008 NSCG Incentive Conditioning Study Sample

2006 Prepaid Incentive?	Cases Eligible for 2008 NSCG Incentive Conditioning Study	2008 Incentive?	Minimum Sample Needed	Extra Cases
Y	1,891	Yes No	698 698	495
N	2,732	Yes No	698 698	1,336
Total	**4,623**		**2,792**	**1,831**

References

Cotterchio, Michelle and Nancy Kreiger (1998). "The Effect of Financial Incentives on Participation in Future Research." *Epidemiology*, 9(2): 216.

Singer, Eleanor, Robert M. Groves and Amy D. Corning (1999). "Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation." *Public Opinion Quarterly*, 63: 251-260

Singer, Eleanor, John Van Hoewyk and Mary P. Maher (1998). "Does the Payment of Incentives Create Expectation Effects?" *Public Opinion Quarterly*, 62: 152-164.