

SF-83-1 SUPPORTING STATEMENT

for

2013

National Survey of College Graduates

CONTENTS

Section	Page
A. JUSTIFICATION	1
1. NECESSITY FOR INFORMATION COLLECTION	7
2. USES OF INFORMATION.....	8
3. CONSIDERATION OF USING IMPROVED TECHNOLOGY.....	11
4. EFFORTS TO IDENTIFY DUPLICATION.....	12
5. EFFORTS TO MINIMIZE BURDEN ON SMALL BUSINESS.....	12
6. CONSEQUENCES OF LESS FREQUENT DATA COLLECTION	12
7. SPECIAL CIRCUMSTANCES.....	13
8. FEDERAL REGISTER ANNOUNCEMENT AND CONSULTATION OUTSIDE THE AGENCY	13
9. PAYMENT OR GIFTS TO RESPONDENTS	16
10. ASSURANCE OF CONFIDENTIALITY.....	17
11. JUSTIFICATION FOR SENSITIVE QUESTIONS	17
12. ESTIMATE OF RESPONDENT BURDEN	17
13. COST BURDEN TO RESPONDENTS	18
14. COST BURDEN TO FEDERAL GOVERNMENT.....	18
15. REASON FOR CHANGE IN BURDEN	18
16. SCHEDULE FOR INFORMATION COLLECTION AND PUBLICATION.....	19
17. DISPLAY OF OMB EXPIRATION DATE.....	19
18. EXCEPTION TO THE CERTIFICATION STATEMENT	19
 B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS	 20
1. RESPONDENT UNIVERSE AND SAMPLING METHODS	20
2. SURVEY METHODOLOGY	21
3. METHODS TO MAXIMIZE RESPONSE	24
4. TESTING OF PROCEDURES.....	26
5. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION	44

APPENDIX A: NSF Act of 1950

APPENDIX B: First Federal Register Announcement

APPENDIX C: 2013 NSCG Coefficient of Variation (CV) Target

APPENDIX D: NSCG Questionnaire Changes

APPENDIX E: Survey Mailing Materials

APPENDIX F: Draft 2013 NSCG Questionnaires

APPENDIX G: Incentive Timing Study Proposal

APPENDIX H: Incentive Conditioning Study Proposal

APPENDIX I: Power Analysis for the 2013 NSCG Methodological Studies

2013 NATIONAL SURVEY OF COLLEGE GRADUATES SUPPORTING STATEMENT

A. JUSTIFICATION

This request is for a three-year revision of the previously approved OMB clearance for the National Survey of College Graduates (NSCG). The NSCG was last conducted in 2010. The OMB clearance for the 2010 NSCG expires July 31, 2013.

The NSCG is one of three principal surveys that provide data for the National Science Foundation's (NSF) 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. Historically, SESTAT has been 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. who had a bachelor's degree or higher in an SEH or SEH-related degree before January 1, 2009, or those who had a bachelor's degree or higher in some other field before January 1, 2009 but having an SEH or SEH-related occupation, including individuals who received degrees only from foreign institutions).

For the 2013 SESTAT survey cycle, the NSRCG has been discontinued in response to the increased coverage of the NSCG offered by its newly implemented sample design. As a result, the 2013 SESTAT will be produced by combining data from the SDR (representing persons in the general U.S. population who have earned a doctorate in SEH from a U.S. institution) and the NSCG (representing all individuals in the U.S. who had a bachelor's degree or higher in a SEH or SEH-related degree before January 1, 2011, or those who had a bachelor's degree or higher in some other field but had an SEH or SEH-related occupation, including individuals who received degrees only from foreign institutions).

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. Historically, the SESTAT surveys have been conducted every two to three years. In the 2010 survey cycle, the NSCG provided information on the U.S. stock of scientists and engineers, 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. For the 2013 survey cycle, with the discontinuation of the NSRCG in response to the expanded coverage of the NSCG, the NSCG and SDR will provide information on the stock of scientists and engineers. Both surveys will also provide information on new graduates with SEH degrees entering the labor force. 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 91% of the population estimate in 2010.

The SESTAT integrated database is the only available source that provides detailed information to support a wide variety of policy and research analyses on science and engineering (S&E)¹ workforce and personnel. To provide complete representation of U.S. S&E workforce 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 past two decades, is closely based on the recommendations of the National Research Council's Committee on National Statistics (CNSTAT) report to NSF.² That report recommended a data collection design based on three surveys, of which one (the NSCG) would be linked to the decennial Census.

Below is the summary of changes in the survey methodology in 2013 from the previous survey year:

1) Continued Implementation of the Rotating Panel Design

Prior to 2010, the new NSCG sample was drawn from the census long form after each decennial census. This long form based sample was then interviewed every two to three years throughout the decade as part of the NSCG sample. With the long form occurring only once every decade, it was not possible to refresh the NSCG sample during the decade. As a result, the long form based NSCG sample suffered from increasing undercoverage of recent graduates and recent immigrants throughout the 1990 and 2000 decades. Furthermore, by only following the S&E population in subsequent survey cycles, the NSCG was not able to provide complete information on people entering or exiting the S&E workforce.

After the 2000 decennial census, the Census Bureau discontinued the long form and introduced the American Community Survey (ACS). In response to this change, NSF commissioned a CNSTAT panel to examine proposed sample design options for the NSCG based on the ACS, as opposed to the long form. The CNSTAT panel issued a 2008 report with recommendations to NSF on the NSCG sample design for the 2010 survey cycle and beyond.³

Using recommendations from this 2008 CNSTAT report, NSF introduced a new rotating panel sample design for the NSCG in the 2010 survey cycle to take advantage of the annual nature of the ACS,. In this rotating panel design, the NSCG selects a new sample every survey cycle from the most recent ACS and follows the cases for four survey cycles. After the fourth cycle, the cases rotate out of the NSCG and are replaced by a newly selected panel of cases from the most recent ACS. When fully implemented, each NSCG survey cycle will include four panels of sample cases with each panel originating from a different ACS year. Through this rotating panel design and the selection of a new sample every NSCG survey

¹ S&E workforce includes the individuals with degrees or occupations in computer and mathematical sciences, life sciences, physical sciences, social sciences, engineering, and health sciences.

² 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.

³ National Research Council, Committee on National Statistics. 2008. *Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics Programs*. Washington: The National Academies Press.

cycle, the NSCG is now able to address the recent graduates and recent immigrants undercoverage that has existed in the past.

The 2010 survey cycle marked the introduction of the use of the ACS sampling frame to select the NSCG sample. In the 2010 NSCG, 65,000 cases were selected from the 2009 ACS. The 2013 survey cycle will continue the implementation of the NSCG rotating panel design by carrying forward the respondents from the 2010 NSCG and by introducing a new panel of sample of 83,000 cases selected from the 2011 ACS. The 83,000 cases to be selected from the 2011 ACS include 65,000 core sample cases and 18,000 cases selected as part of a young college graduates oversample. The 2010 NSCG respondents carried forward will be referred to as the old cohort cases and the new sample cases selected from the 2011 ACS will be referred to as the new cohort cases.

Full implementation of the NSCG four-panel rotating panel design is expected to occur in the 2017 survey cycle. Once the rotating panel design is fully implemented, each survey cycle will see the addition of approximately 32,500 cases from the most recent ACS to offset the rotating out of the oldest NSCG panel. In addition, each rotating panel will include an oversample of young graduates to allow the NSCG to enable more detailed evaluation of the young graduates population in response to the NSF decision to discontinue the NSRCG after the 2010 survey cycle.

2) Discontinuation of the NSRCG

In the 1989 CNSTAT report that led to the establishment of the current SESTAT design, the CNSTAT panel recommended that NSF implement a biennial survey to address the undercoverage of new graduates that exists in the long form based design of the NSCG. This recommendation led to the creation of the NSRCG. As a result, throughout the 1990 and 2000 decades, the NSRCG provided SESTAT with coverage of recent bachelor's or master's degree recipients in SEH degree fields from U.S. institutions.

In the 2010 survey cycle, the NSCG began selecting sample from the ACS and, through the rotating panel design, the NSCG was now able to provide coverage of the recent graduates population throughout the decade. With this increased coverage available through the NSCG, the NSF conducted an evaluation to investigate the possibility of a SESTAT design change that would include discontinuing the NSRCG and using the NSCG, with an expanded sample of young graduates, to provide coverage of this recent graduates population. As part of this evaluation, the NSF completed the following investigation steps:

- Conducted extensive outreach to determine the impact this design change would have on the S&E community. The audience for the outreach efforts included, but was not limited to, the American Association for the Advancement of Sciences, the Association for Institutional Research, the Association of American Medical Colleges, the Association of American Universities, the Committee on Equal Opportunities in Science and Engineering, the Council of Graduate Schools, the NCSES Human Resources Experts Panel, the National Center for Education Statistics, the Census Bureau, and numerous divisions within NSF.

- Compared the precision level of recent graduate estimates from the NSRCG and from the ACS-based design of the NSCG and examined possible oversampling strategies to increase the NSCG sample for improved precision of recent graduates estimates.
- Examined the cost implications of this design change.

After completing this evaluation, the NSF created a summary report and presented the findings at a CNSTAT dissemination meeting. Findings from the evaluation included the following.

- Through our outreach discussions with the S&E community, the NSF found few instances of the NSRCG being used as a stand-alone file for analytical purposes. As a result, we concluded it is the ability to provide coverage of the recent college graduates population within the context of SESTAT that makes the NSRCG valuable to analysts.
- On the topic of the recent graduates estimate comparison, the NSF concluded that the NSRCG provided better precision for estimates of graduates within the previous two academic years, but that the NSCG with a young graduates oversample had the potential to provide better precision for analysts working with an expanded definition of “recent” that includes those earning degrees in the previous five academic years.
- When we asked a subset of data users whether they preferred defining “recent” as “within the previous two academic years” or as “within the previous five academic years,” most reported no preference or preferred the expanded definition based on five academic years.
- When the costs for the current design and the proposed design were compared, it was determined that the discontinuation of the NSRCG and the expansion of the young graduates sample in the NSCG had the potential to reduce the overall cost of the SESTAT program by three million dollars every biennial survey cycle.

After reviewing our evaluation results and carefully considering the feedback received from the extensive outreach efforts with the S&E community, NCSES decided to discontinue the NSRCG after the 2010 survey cycle. A major impetus for this decision is that the NSRCG is no longer needed to fill the coverage gaps of SESTAT. Instead, the NSCG, through the use of the ACS-based sampling frame and its rotating panel design, provides on-going coverage of the recent college graduates population. Other factors considered in this decision were the limited use of the NSRCG as a standalone data file and the cost savings associated with discontinuing the NSRCG and with simplifying the SESTAT integration processes. Finally, to enable analysts to evaluate the recent college graduates population (using an expanded definition of “recent”), NCSES plans to expand the sample of young college graduates in the 2013 NSCG.

3) Young College Graduates Oversample in the NSCG

To allow continued analysis of the recent bachelor's and master's SEH degree recipients that had been available from the NSRCG, the 2013 NSCG will include an oversample of 18,000 young college graduates selected from the 2011 ACS. This oversample will not enable the same level of precision for estimates of bachelor's and master's graduates within the previous two academic years as had been available through the NSRCG, but the NSCG oversample does have the potential to provide better precision for estimates of bachelor's and master's degree recipients from the previous five years.

While it is not possible to precisely oversample the recent bachelor's and master's SEH degree recipients as was done in the NSRCG, the information available on the ACS-based sampling frame does allow the ability to develop an oversampling scheme that can increase the number of recent bachelor's and master's SEH degree recipients within the NSCG responding sample. Questionnaire items included on the ACS such as educational attainment, bachelor's field of degree, age, and recent school enrollment may enable the development of an oversampling scheme that results in a substantial increase in the number of recent bachelor's and master's S&E degree recipients in the NSCG responding sample.

To determine what information available on the ACS best predicts the likelihood of a case reporting a recent college degree on the NSCG questionnaire, the Census Bureau and the NSF conducted analysis that used the ACS-based sampling frame information in combination with the 2010 NSCG response information. The goal of the analysis was to identify strong predictors for recent degree likelihood. These predictors could then be used to develop an oversampling scheme for the NSCG that would result in an increased number of recent college graduates in the NSCG responding sample. The ACS questionnaire items used in this analysis to determine predictors of recent degree likelihood were the following:

- Educational attainment⁴
- Bachelor's field of degree
- Age
- Recent school enrollment indicator⁵

As part of the analysis, we examined multiple combinations of the age and recent school enrollment indicator in an effort to determine oversampling criteria that best predicts recent

⁴ This variable was used to enable the analysis to focus on increasing the sample for only the recent bachelor's and master's degree recipients in the NSCG. This decision to focus only on bachelor's and master's degree recipients was made to align with the NSRCG coverage. No attempts were made to increase the sample of recent doctoral degree recipients in the NSCG as part of this evaluation.

⁵ This ACS questionnaire item asks respondents the following question: *At any time in the last 3 months, has this person attended school or college?* If the respondent answers "Yes" to this question, the skip pattern leads to a follow-up questions that asks them to provide the grade or level at which they were attending. The options provided are nursery school, preschool; Kindergarten; Grade 1 through 12; College undergraduate years (freshman to senior); and Graduate or professional school beyond a bachelor's degree.

SEH degree likelihood within the NSCG responding sample. From this analysis, we decided to use the following criteria to define the cases eligible for the young graduates oversample within the 2013 NSCG.

- ACS sample cases with a bachelor's degree who are ages 28 or less and are educated or employed in an SEH field
- ACS sample cases with a master's degree who are ages 32 or less and are educated or employed in an SEH field

4) Web First Data Collection Strategy

The 2010 NSCG survey cycle marked the introduction of a web data collection mode to compliment the mail questionnaire and computer-assisted telephone interviewing (CATI) options that had existed in previous survey cycles. Since 2010 was the first time the web mode was available as a response option in the NSCG, the NSF decided to take a cautious approach with rolling out the web mode to potential respondents. In the 2010 NSCG, the default data collection path (i.e., the data collection path offered to most cases) included an initial mailing that offered the choice for responding by web or mail. This default data collection path was based on our past practice of using mail as an initial mode, but included the web mode under the assumption that a web response option is apt to be appealing to the highly web-literate college-educated population that is the focus of the NSCG.

To better assess the potential of the web mode for future survey cycles, the 2010 NSCG included a mode effects experiment. In this experiment, a subset of the 2010 NSCG sample cases selected from the 2009 ACS were randomly assigned to three treatment groups: mail first, web first, and CATI first. The cases stayed in their assigned mode (mail, web, or CATI) for the first eight weeks of data collection and then were sequentially offered the other modes. Through this experiment, we found that the web first approach could produce final response rates that exceeded or were not statistically different from the final response rates for the mail first and CATI first approaches. In addition, by conducting a detailed evaluation of the data collection costs, we determined that the web first approach achieved these impressive response results at a much lower cost per respondent (approximately \$50 per respondent in the web first approach versus \$65 in the mail first approach and \$75 in the CATI first approach). Finally, the research showed that the majority of respondents tended to respond in the initially offered mode. This finding held across all three treatment groups. So, if there is a desire to have respondents complete the NSCG by web, using a web first approach will increase the likelihood of a web response. Given the positive findings from the 2010 NSCG mode effects experiment, the 2013 NSCG will use the web first data collection approach as its default data collection path.

It should be noted that while the web first data collection approach will be the default data collection path used in the 2013 NSCG, exceptions to this default path will be made to honor mode preference for certain cases in the 2013 NSCG data collection effort. The 2010 NSCG questionnaire included an item that asked all respondents to let us know how they “would like to complete the survey in future rounds.” The 2010 NSCG respondents were given the options of a questionnaire sent in the mail, a web questionnaire on the internet, a telephone

interview, or no preference. For the 2013 NSCG survey cycle, we will make exceptions to the default data collection path for cases that stated a preference to complete the survey by mail or telephone. For the cases that reported a mail questionnaire preference, a paper questionnaire will be sent to respondents at the beginning of the 2013 NSCG data collection effort. Similarly, for the cases that reported a telephone interview preference, telephone calls asking the sample case to complete the survey will be made during the first seven weeks of data collection when the default path cases will only receive mailings encouraging response by web.

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*), NSF, 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 careers, training, 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 the 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.

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 American Institutes of Research used the SESTAT data in the evaluation of the NSF's Alliance for Graduate Education and the Professoriate Program; the Commission on Professionals in Science and Technology regularly publishes data from SESTAT in their STEM Trends publications; the General Accounting Office used the SESTAT data to issue a report on education and disability. The Federal Reserve Bank of St. Louis used the SESTAT data to examine the pathway from

Community College to a Bachelor's Degree and Beyond; and many 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 public-use files. The NSCG panel data from all the 1990s and 2000s cycles are also available as a component of the SESTAT database for each survey year (1993, 1995, 1997, 1999, 2003, 2006 and 2008), which are available as SESTAT public-use files. The 2010 NSCG data are in the final stages of data review and will be available later this year as a standalone file.

The SESTAT data were used extensively in the latest versions of the congressionally mandated biennial reports *Science and Engineering Indicators, 2012* and *Women, Minorities and Persons with Disabilities in Science and Engineering, 2011*. In addition, the *Women, Minorities and Persons with Disabilities in Science and Engineering, 2013*, set for release in early 2013 will also use SESTAT data.

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

- *2003 College Graduates in the U.S. Workforce: A Profile*, December 2005
- *What Do People Do After Earning a Science and Engineering Bachelor's Degree?* June 2006
- *Why Did They Come to the United States? A Profile of Immigrant Scientists and Engineers*, June 2007
- *Unemployment Rate of U.S. Scientists and Engineers Drops to Record Low 2.5% in 2006*, March 2008
- *Diversity in Science and Engineering Employment in Industry*, March 2012

All NSF Publications can be accessed on the National Center for Science and Engineering Statistics (NCSES) 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 internet. The SESTAT on-line system allows internet users to create customized data tabulations with a user-specified subject area. Additionally, the NSCG and SESTAT public-use files are available for download through the SESTAT web page 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 meetings of the Association for Institutional Research or the American Educational Research Association.

Since 2007, NSF has distributed over 200 copies of the more than decade-old 1993 NSCG public-use data set to researchers in government, academia, and professional societies. In addition, over 700 copies of the 2003 NSCG public-use files have been requested since 2007. In spite of the age of the data, the 1993 and 2003 NSCG data continue to be heavily used because they are the only data sets 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 between 1993 and 2008 included college degree holders who earned their degrees outside of the United States and who were residing here at the time of the previous census.

There are currently 20 licensed users for the SESTAT integrated database micro data files under a licensing agreement with NCSES. Over 2,400 users have downloaded the SESTAT public-use files since 2007. As previously noted, over half of the records in the SESTAT file come from the NSCG.

Some of the research from the public-use NSCG data and the SESTAT restricted data licensees resulted in papers such as:

- *Why Don't Women Patent?*, National Bureau of Economic Research, 2012
- *Findings from an Examination of the Labor Force Participation of College-Educated Immigrants in the United States*, Department of Education, 2012
- *Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors*, New York University, 2011
- *Earning Trajectories of Highly Educated Immigrants: Does Place of Education Matter?*, Cornell University, 2011
- *Which Immigrants are Most Innovative and Entrepreneurial? Distinctions by Entry Visa*, National Bureau of Economic Research, 2011
- *Labor Market Penalties for Foreign Degrees Among College Educated Immigrants*, University of Minnesota, 2010
- *Do Teachers have Education Degrees? Matching Fields of Study to Popular Occupations of Bachelor's Degree Graduates*, Indiana University, 2010
- *Why Do Women Leave Science and Engineering?*, National Bureau of Economic Research, 2010
- *Functional Impairment and the Choice of College Major*, University of South Florida, 2010
- *How Much Does Immigration Boost Innovation?*, McGill University, 2010
- *Increasing Time to Baccalaureate Degree in the United States*, National Bureau of Economic Research, 2010
- *Higher Education and Disability: Education Needs a Coordinated Approach to Improve Its Assistance to Schools in Supporting Students*, GAO Report, 2009

- *Diversifying Science and Engineering Faculties: Intersections of Race, Ethnicity, and Gender*, Georgia Institute of Technology, 2010
- *Earnings of a Lifetime: Comparing Women and Men with College and Graduate Degrees*, Indiana University Kelley School of Business, 2009
- *Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors*, Harvard University, 2009
- *The Small Firm Effect and the Entrepreneurial Spawning of Scientists and Engineers*, Washington University St Louis, 2009
- *From Community College to a Bachelor's Degree and Beyond: How Smooth Is the Road?*, Federal Reserve Bank of St. Louis, 2009
- *Double Your Major, Double Your Return?*, St. Lawrence University, 2008
- *Gender Wage Disparities Among the Highly Educated*, University of Chicago, 2008

3. CONSIDERATION OF USING IMPROVED TECHNOLOGY

The Census Bureau will collect the 2013 NSCG data under an interagency agreement, using a multi-mode approach, with a web invitation letter mailed to sample persons asking them to complete the survey on the Internet. Nonrespondents will be followed up using a paper questionnaire mailing and CATI. The change to this web first data collection approach is based on findings from the 2010 NSCG mode effects experiment that showed both a response and cost advantage to using the web first approach. The NSCG web survey will be developed using Census Centurion, which is a secure web-based application programmed to meet the stringent Census Bureau data security requirements. The web survey will take advantage of the computer-assisted interviewing system that allows for probes both for invalid or inconsistent responses and for missing responses to a few question items critical for a complete interview.

Because the sample contact information will be at least a year old for most sample members 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 non-mailable 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 2013 NSCG will investigate several facets of adaptive design in an effort to attain high-quality survey estimates in less time and at less cost than traditionally executed survey operations. First, the Census Bureau will implement daily processing (editing, imputation, weighting) of the response data throughout the data collection period. In prior survey cycles, the processing occurred toward the end or after the data collection period. The daily processing approach is expected to reduce the overall time from the beginning of data collection until the

final delivery of data and estimates. In addition to operational efficiencies, daily processing will allow the NSCG survey team to monitor several quality measures throughout data collection, including R-indicators, benchmarking, stability of estimates, and response propensities by mode.

In addition to the change to a daily processing approach, adaptive design techniques will be directly employed in a mode-switching experiment where data quality measures will be examined on a weekly basis, and cases will be switched between data collection modes, or put on hold entirely. This experiment is an attempt to allocate resources more efficiently in order to maximize survey quality while minimizing wasted funds and effort. The specifics of this experiment are provided in Section B.4.

4. EFFORTS TO IDENTIFY DUPLICATION

Duplication, in the sense of similar data collections, 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 provides occupational estimates but does not collect information on degree field for higher education degrees. The ACS collects the field of bachelor's degrees but does not collect detailed information on education history, work activities, and employment characteristics as the NSCG does, nor is the ACS longitudinal in nature.

Since there is overlap in the target populations of the NSCG and the SDR, efforts were taken in 2010 to identify any cases selected for sample in both surveys. Any duplicates identified were removed from the NSCG data collection efforts and were only prompted to respond to the SDR. At the conclusion of the SDR data collection effort, final interview disposition and any response information was copied to the NSCG data file to allow the information for these duplicate cases to exist in both surveys. A similar deduplication process between the NSCG and SDR is tentatively planned for the 2013 survey cycle.

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 has been conducted and coordinated with the NSRCG and the SDR from 1993 through 2010 and will be conducted and coordinated with the SDR in 2013. 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 June 1, 2012 (See Appendix B.) NSF received no public comment in response to the announcement as of the close date of July 31, 2012.

Consultations Outside the Agency

The National Center for Science and Engineering Statistics (NCSES)⁶ within the NSF has responsibility for the SESTAT surveys. In the early 1990s, NCSES 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 past two decades.

As the SESTAT survey system entered the 21st century, NCSES 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 analysts. To accomplish this goal, NCSES carefully planned and engaged in a series of formal and informal evaluations and assessments of

⁶ Prior to 2011, the National Center for Science and Engineering Statistics was known as the Division for Science and Resources Statistics (SRS). While many of the activities discussed in this section occurred under the SRS name, we will use the NCSES name throughout this section for simplicity. Both names (SRS and NCSES) refer to the same organizational unit within NSF.

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, NCSES 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 has continued to take place through a series of meetings and workshops on various issues related to the SESTAT redesign and survey methodology since 2008.

For the 2010 survey round:

- NCSES worked with the U.S. Census Bureau, OMB, and other Federal agencies to add a field of degree (FOD) question to the ACS, to enable more precise sampling for future NSCG surveys. As a part of this activity, NCSES worked with the Census Bureau on a methods test to test various versions of a FOD question.
- NCSES commissioned the Committee on National Statistics (CNSTAT) of the National Research Council (NRC) to examine proposed sample design options for the NSCG based on the ACS, as opposed to the long form of the Decennial Census. The CNSTAT committee held a two-day workshop on this topic, and issued a report with recommendations to NSF on the 2010 and beyond NSCG sample design. The recommendations formed the basis for the 2010 NSCG design.⁷
- NCSES coordinated with OMB on wording for the collection of data on functional disability question items in the SESTAT surveys to increase consistency across the Federal statistical agencies in surveys with such questions. As a result, a new category on cognitive disability, taken from the ACS, was added to all three SESTAT surveys in 2010, and the introductory sentence was revised to refer to difficulties with specific functional limitations.

For the 2013 survey round:

- NCSES evaluated a possible SESTAT redesign to improve the timeliness, quality, and efficiency of the surveys combined to form the Scientists and Engineers Statistical Data System (SESTAT) while, if possible, reducing overall survey costs. The evaluation examined the potential impact on the science and engineering (S&E) community, on the precision of SESTAT estimates, on data usage, and on survey cost. The decision to examine SESTAT was partially motivated by a 2008 CNSTAT recommendation for NSF

⁷ National Research Council, Committee on National Statistics. 2008. *Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics Programs*. Washington: The National Academies Press.

to “use the opportunity afforded by the introduction of the ACS as a sampling frame to reconsider the design of the SESTAT Program and the content of its component surveys.”

To obtain feedback from the S&E community, NCSES conducted extensive outreach efforts with a broad audience including, but not limited to, the American Association for the Advancement of Science; Association for Institutional Research; Association of American Medical Colleges; Association of American Universities; Committee on Equal Opportunities in Science and Engineering (CEOSE); Council of Graduate Schools; NCSES Human Resources Experts Panel; National Center for Education Statistics; the Census Bureau; and, within NSF.

- After reviewing the evaluation results and carefully considering the feedback received from the outreach efforts with the S&E community, NCSES decided to discontinue the NSRCG after the 2010 survey cycle. A major impetus for this decision was that the NSRCG is no longer needed to fill the coverage gaps of SESTAT. Instead, the NSCG, through the use of the ACS, provides on-going coverage of the recent college graduates population. Other factors considered in this decision were the limited use of the NSRCG as a standalone data file and the cost savings associated with discontinuing the NSRCG and with simplifying the SESTAT integration processes. NCSES plans to expand the sample of young college graduates in the NSCG beginning with the 2013 survey. As part of the decision-making process related to this proposed design change, NCSES held a debriefing meeting with the 2008 CNSTAT panel members to discuss the evaluation findings. At the meeting, NCSES received no objections to its proposal to discontinue the NSRCG and to expand the sample of young graduates in the NSCG.

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, NCSES has engaged in the following activities.

For the 2006, 2008, 2010, and 2013 survey rounds:

- NCSES has convened a Human Resources Experts Panel (HREP) in order to help improve data collection on the S&E workforce through review of the S&E personnel surveys and to promote use of the data for research and policy analysis purposes. HREP accomplished 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 policy 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 NCSES human resources surveys. The panel includes 15 members who represented the sciences, academia, business/industry, government, researchers and policy makers. Seven meetings have been held since the panel was 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. NCSES has funded three workshops where 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, NCSES funded a workshop at NSF that brought in experts on database matching. NCSES is currently engaged in an activity that will enable the matching of some SESTAT data, specifically the Survey of Doctorate Recipients (SDR) data, to various patent and publication databases.
- Through a grant to the Association for Institutional Research (AIR), NCSES staff recorded two webinars on the SESTAT website and data tool to encourage broader use of the data.
- ASA/AAPOR invited an NCSES analyst to present a webinar on science and technology human resources surveys, data and indicators; the SESTAT data are the source for all of the major indicators and trends on this workforce.

9. PAYMENT OR GIFTS TO RESPONDENTS

Motivated by the findings from the late stage incentive included in the 2010 NSCG in combination with the desire to obtain a better understanding of optimal incentive usage in data collection efforts, the NSF is considering two monetary incentive experiments to examine potential nonresponse bias in the 2013 NSCG: an incentive timing study and an incentive conditioning study. The incentive timing study will examine the impact that the timing of when an incentive is offered has on response rate, sample representation, and cost. The incentive conditioning study will examine the impact that a previous incentive has on a sample case's propensity to respond in a subsequent survey cycle.

The incentive in both studies will be a \$30 prepaid debit card incentive that is similar to the debit card incentive used in the 2010 NSCG survey cycle. These debit cards will have a six month usage period at which time the cards will expire and the unused funds will be returned to Census and NSF (minus the predetermined per card fee).

Preliminary design information for both incentive studies are discussed later in this document (Section B.4.) and in the appendices.

10. ASSURANCE OF CONFIDENTIALITY

NSF and the Census Bureau are committed to protecting the confidentiality of all survey respondents. The NSCG data will be collected in conformance with the Privacy Act of 1974, the NSF Act of 1950, as amended, and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002. The Census Bureau is conducting the NSCG under the authority of Title 13 and 15, United States Code, Section 8 and 1525, respectively.

As explained in Section B.1, there are three components of the 2013 NSCG sample design. The first one is the 2010 NSCG respondents from the 2009 ACS, the second is “NSRCG panel” respondents subsampled from the 2010 NSRCG, and the third is based on respondents to the 2011 ACS.

The statement on the questionnaire cover will cite the appropriate 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. The cover letters will include additional statements in the Frequently Asked Questions section 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 web phase and 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 “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 144,000 sample persons by web, mail or computer-assisted interviewing as part of the 2013 NSCG data collection. Based on experience administering the NSCG interviews, the questionnaire takes an average of 25 minutes to complete. An overall response rate of about 80 percent is estimated from the 83,000 new cohort

sample, and an overall response rate of about 90 percent from the 61,000 old cohort sample. Based on an estimate of approximately 121,300 completed cases, the total burden hours for the 2013 NSCG data collection are 50,542. The total cost to respondents for the 50,542 burden hours is estimated to be \$1,579,437. This estimate is based on an estimated median annual salary of \$65,000 per NSCG employed respondent. Assuming a 40-hour workweek and a 52-week salary, this annual salary translates to an hourly salary of \$31.25. Salary estimates were obtained using data from the 2010 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 2013 NSCG is approximately \$14.6 million, which includes survey cycle costs, and NSF staff costs to provide oversight and coordination with the other SESTAT survey. The estimate for survey cycle costs is approximately \$14 million, 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 \$562,500 (based on \$150,000 annual salary of 1.5 FTE for 2.5 years).

15. REASON FOR CHANGE IN BURDEN

In the past, after each decennial census, a new sample was drawn from the census long form, and that sample was followed until the end of the decade. This was done in 1982, 1993 and 2003. The first survey in the decade has been much more expensive and burdensome than the following ones because of the larger sample size required to identify the S&E personnel in the U.S.

Through the rotating panel design established for the NSCG's use of the ACS-based sampling frame, the NSCG sample size will stay more consistent throughout the decade as opposed to the once per decade sample size increase experienced using the long form based sampling frame. Under the current sample design plans, the average biennial sample size for the NSCG rotating panel design will be near 110,000 cases. If we consider only the 2010 NSCG respondents (47,000 cases) and the core sample selected from the 2011 ACS (65,000 cases), the sample size is near the proposed 110,000 per cycle sample size. It is the discontinuation of the NSRCG, the decision to follow the 2010 NSRCG cases as part of the NSCG, and the decision to oversample young graduates in the NSCG that results in the burden change between the 2010 NSCG and the 2013 NSCG. The 30,00 sample size resulting from the inclusion of the 2010 NSRCG cases

(12,000 cases) and the young graduates oversample (18,000 cases) explains the majority of the difference in burden hours between the 2010 NSCG (34,792 burden hours) and the 2013 NSCG (50,542 burden hours).

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 2013 data collection and publication is currently estimated as follows:

Data Collection	February 2013 – July 2013
Coding and Data Editing	February 2013 – January 2014
Final Edited/Weighted/Imputed Data File	February 2014
SESTAT Info Brief	Late Spring 2014
SESTAT Detailed Statistical Tables	Summer 2014
SESTAT Integrated Public Use Data File	Summer/Fall 2014

17. DISPLAY OF OMB EXPIRATION DATE

The OMB Expiration Date will be displayed on the 2013 NSCG questionnaires.

18. EXCEPTION TO THE CERTIFICATION STATEMENT

Not Applicable.