

SF-83-1 SUPPORTING STATEMENT

For

2013 Survey of Doctorate Recipients

TABLE OF CONTENTS

A. Justification	3
1. Necessity for Information Collection	4
2. Uses of Information	5
3. Consideration of Using Improved Technology.....	7
4. Efforts to Identify Duplication	8
5. Efforts to Minimize Burden on Small Business.....	8
6. Consequences of Less Frequent Data Collection	8
7. Special Circumstances	8
8. Federal Register Announcement and Consultations Outside the Agency.....	8
9. Payment or Gifts to Respondents	10
10. Assurance of Confidentiality	12
11. Justification for Sensitive Questions	13
12. Estimate of Respondent Burden	13
13. Cost Burden to Respondents.....	13
14. Cost Burden to the Federal Government.....	14
15. Reason for Change in Burden	14
16. Schedule for Information Collection and Publication	14
17. Display of OMB Expiration Date.....	14
18. Exception to the Certification Statement	14
B. Collection of Information Employing Statistical Methods	15
1. Respondent Universe and Sampling Methods	15
2. Statistical Procedures.....	16
3. Methods to Maximize Response.....	17
4. Testing of Procedures	18
5. Contacts for Statistical Aspects of Data Collection.....	19

LIST OF ATTACHMENTS

Attachment 1: National Science Foundation Act of 1950

Attachment 2: First Federal Register Notice for 2013 SDR

Attachment 3: Proposed 2013 SDR Mailing Materials

Attachment 4: 2010 SDR Questionnaire

Attachment 5: Changes in the 2013 SDR Methodology

Attachment 6: Changes in the 2013 SDR Questionnaire

Attachment 7: Items Asked Only in the CATI and Web Instruments

2013 SURVEY OF DOCTORATE RECIPIENTS

Supporting Statement

A. Justification

This request is for a three-year revision of the previously approved OMB clearance for the Survey of Doctorate Recipients (SDR). The SDR was last conducted in 2010 and the OMB clearance for the 2010 SDR expires July 31, 2013 (OMB No 3145-0020).

The SDR provides 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. SESTAT is produced by combining data from the 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) with data from the National Survey of College Graduates (representing all individuals in the U.S. with a bachelor's or higher degree in an SEH or SEH-related field, or those with a bachelor's or higher degree in some other field, but having an SEH or SEH-related occupation, including individuals who received degrees only from foreign institutions).

The SESTAT integrated database derived from these surveys contains data on the demographic, educational, and employment characteristics of college-educated scientists and engineers in the United States. These surveys are usually conducted every two years. The primary purpose of the SDR is to provide information on doctoral scientists and engineers who were awarded degrees from U.S. institutions and reside in the U.S. It is comprised of two components: 1) a longitudinal panel that tracks doctorate recipients throughout their careers until age 76, and 2) a new cohort component that adds new doctorate recipients after they receive their degree. The panel portion of the SDR provides information on the experienced stock of doctorate recipients, while the new sample in the SDR provides important data on the early career experiences of new doctorate recipients with SEH degrees entering the labor force.

In addition, since 2003 and continuing with the 2006, 2008, and 2010 SDR, the NSF tested and reaffirmed the feasibility of developing a complimentary international panel study of U.S. trained doctorate recipients. This sub-sample was comprised primarily of non-U.S. citizens who emigrated after degree award. The 2013 SDR will represent both a National Survey of Doctorate Recipients (NSDR) to be included in the SESTAT, and a smaller International Survey of Doctorate Recipients (ISDR), which will include U.S. citizens as well as non-U.S. citizens living outside the U.S. Currently, 33% of U.S. SEH doctorates are awarded to temporary visa holders and nearly 27% of them plan to leave the U.S. upon graduation. The 2013 ISDR will yield new information about the educational and demographic characteristics of U.S. trained SEH doctorate recipients living and working abroad on the reference date, 1 February 2013.

The SDR, as part of the SESTAT data system, is the only available source that provides detailed information at the doctorate level to support a wide variety of policy and research analyses on science, engineering and health (SEH) labor force issues. To provide complete representation of U.S. scientists and engineers at all degree levels, SESTAT was designed as a unified database that integrates information from 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 (NRC) Committee on National Statistics (CNSTAT) report to NSF.¹ That report

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.

recommended a data collection design based on three surveys (NSCG, SDR, and the National Survey of Recent College Graduates (NSRCG)). Now that the NSCG sample is based on the American Community Survey (ACS) rather than the decennial Census, the NSRCG has been deemed unnecessary and SESTAT can be based on two surveys (NSCG and SDR).

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 Attachment 1 – National Science Foundation Act of 1950.)

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, 1st Session, Report No. 92-204).

Subsequently, the NSF established and continues to maintain the SESTAT system, 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 the SDR is a major part, provides much of the information to meet this mandate.

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

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.

In addition, 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. Similarly, ad hoc committees convened by the National Research Council of the National Academies (Advisors to the Nation on Science, Engineering, and Medicine) have used SDR and SESTAT data in Committee reports such as the Committee on Gender Differences in Careers of Science, Engineering, and Mathematics Faculty's 2009 report "*Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty.*"

2. Uses of Information

The time-series data produced by the SDR on the demographic, employment, and other characteristics of the Nation's SEH doctoral scientists and engineers have been used extensively in the policy and planning activities of the Foundation and the National Institutes of Health. The SDR data are used in assessing the quality and supply of the Nation's S&E personnel resources for educational institutions, private industry, and professional organizations as well as federal, state, and local governments. Other federal agencies, such as the Department of Commerce, USDA, DOE, and NASA, as well as state agencies request and make use of the SDR data for a variety of informational purposes.

SDR data are also an integral part of SESTAT. Researchers, policymakers, and others use information from the SESTAT database to answer questions about the number, employment, education, and characteristics of the S&E workforce. Because SESTAT 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 postdoctoral appointments, 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.

Findings from the 2013 SDR will enable the NSF to continue monitoring employment patterns of recent SEH doctorate recipients, as well as more experienced doctorate recipients in the labor market. The SDR data on the state of SEH doctorates are used for presentations to the National Science Board. Within the Foundation, SDR data are used in the evaluation and development of programs in the Education and Human Resources (EHR) Directorate, and analysis of employment pathways by several research directorates.

The SDR provides data on the educational training, work experience, and career development of persons holding SEH doctorates from U.S. institutions. Without this information, those at the NSF, along with researchers and policymakers, would be less informed when carrying out their responsibilities. The SDR data are made available through published reports, the SESTAT on-line data system, public use files and restricted licenses.

Some recent examples of NSF Publications that used the SDR data (all NSF publications can be accessed on the NCSES website at <http://www.nsf.gov/statistics>) include:

Congressionally mandated reports –

- *Science & Engineering Indicators*
- *Women, Minorities and Persons with Disabilities in Science and Engineering*

Other NSF publications –

- Biennial report series: *Characteristics of Doctoral Scientists and Engineers in the United States*
- Annual report series: *Science and Engineering State Profiles*
- Racial and Ethnic Diversity among U.S.-Educated Science, Engineering, and Health Doctorate Recipients: Methods of Reporting Diversity (January 2012)
- Academic Institutions of Minority Faculty with Science, Engineering, and Health Doctorates (October 2011)
- Unemployment among Doctoral Scientists and Engineers Remained Below the National Average in 2008 (January 2011)
- The End of Mandatory Retirement for Doctoral Scientists and Engineers in Postsecondary Institutions: Retirement Patterns 10 Years Later (December 2010)

Data Dissemination and Access:

The SDR data have been incorporated in the SESTAT on-line data system for each survey cycle since 1993 and are available as a component of the SESTAT public-use data files, or as separate stand-alone public-use files, or as restricted use files licensed by NSF. The SESTAT on-line system allows Internet users to create customized data tabulations in subject areas of their interest. The SESTAT Home Page can be accessed at <http://www.nsf.gov/statistics/sestat>.

Results from the SESTAT integrated data and SDR data are routinely presented at conferences and professional meetings, such as the annual meetings of the Association for Institutional Research, the American Association for Public Opinion Research, and the American Educational Research Association.

Since 2007, NSF has distributed nearly 900 copies of SDR public-use files (2001, 2003, 2006, and 2008 survey cycles), as well as over 2,400 copies of the SESTAT public-use files, of which SDR is a part (1993-2008 survey cycles). There are currently 44 licenses for use of the SDR; there are also 20 licenses for the SESTAT data, which includes the SDR. Additional licensing requests for the SDR are pending review and approval by NSF.

Recent examples of use of the SDR data include the following:

Selected Presentations:

- *Migration Patterns of U.S. Trained Doctorate Holders (A Longitudinal Study)*, Joint Statistical Meetings, August 2012.
- *Development of the Sample Design for the International Survey of Doctorate Recipients*, Joint Statistical Meetings, August 2012.
- *Coping with Missing Data: Assessing Methods for Logically Assigning Race and Ethnicity*, American Association for Public Opinion Research, May 2012
- *Science and Engineering Doctorate Recipients as Adjunct Faculty: New Findings from the Survey of Doctorate Recipients*, American Educational Research Association, April 2012.
- *An investment in Goodwill or Encouraging Delays? Examining the Effects of Incentives in a Longitudinal Study*, Federal Committee on Statistical Methodology Annual Meeting, January 2012.

Selected Citations of SDR data in other sources:

- *Education and Career Outcomes for Women of Color in Academia*, National Academies'

Conference Seeking Solutions: Maximizing American Talent by Advancing Women of Color in Academia, 2012.

- *Keeping Women in the Science Pipeline*, The Annals of the American Academy of Political and Social Science, 2011.
- *Psychology's Researchers & Educators: What's on our workforce horizon?* American Psychological Association, Center for Workforce Studies, APA Annual Conference, 2010.
- *Stay Rates of Foreign Doctorate Recipients from U.S. Universities, 2006*, Science and Engineering Education Program of Oak Ridge Institute for Science and Education, 2009.
- *The Impact of Information Technology on Scientists' Productivity, Quality, and Collaboration Patterns*, National Bureau of Economic Research, 2009.
- *Who's Patenting in the University? Evidence from the Survey of Doctorate Recipients*, Economics of Innovation and New Technology, 2009
- *Ethnic and Technical Clustering: Native-Born Americans versus Foreign S&E Graduates*, International Studies in Entrepreneurship, 2008.
- *Negative Effects of University Patenting: Myths and Grounded Evidence*, Scientometrics, 2008.
- *Problems in the Pipeline: Gender, Marriage, and Fertility in the Ivory Tower*, Journal of Higher Education, 2008.

3. Consideration of Using Improved Technology

The 2013 SDR will collect data using three different modes of data collection: 1) self-administered online surveys; 2) self-administered paper mail questionnaires; and 3) computer-assisted telephone interviews (CATI). Until the 2003 survey cycle, SDR data were collected by first mailing paper questionnaires to sample persons and then following up the nonrespondents with CATI. The tri-mode data collection effort including mail, CATI, and Web was tested in the 2003 SDR and has been fully implemented in all of the rounds since (i.e., 2006, 2008, and 2010 SDR). The 2013 survey cycle will continue this fully implemented tri-mode data collection protocol.

In 2008, over 57 percent of sample members completed a Web survey and in 2010, that number rose to 63 percent. The 2013 SDR will honor mode preferences reported in the 2010 SDR but also emphasize the efficiency of completing via the Web. NSF expects that 65 percent or more of the 2013 survey responses will be in the Web mode.

The 2013 data collection effort will use a comprehensive computerized case management system to track data capture across the three modes (Web, mail, CATI). Optical scanning will be used to capture the digital images of the mail questionnaire after keying. The images will be stored in a database for archival purposes.

4. Efforts to Identify Duplication

Duplication does not exist. No other data collection is based on a probability sample of the U.S. trained doctoral population in science, engineering and health fields living in the United States and more recently since the 2003 SDR, living abroad (as part of the ISDR). Data from the Current Population Survey and the American Community Survey provide occupational estimates and only estimates of degree field earned at the bachelor's level. The 2013 SDR is necessary to obtain trend data on continuing education and career paths of U.S. trained doctorate holders in science, engineering and health fields as well as data that reflect trends in employment patterns. There is no similar information available on this highly trained population that may be used, modified, or made comparable to the SDR.

5. Efforts to Minimize Burden on Small Business

Not applicable. The SDR collects information from individuals only.

6. Consequences of Less Frequent Data Collection

Follow-up surveys every two years on the same sampled persons are necessary to track changes in the SEH workforce due to large movements in and out of SEH occupations over both business cycles and life cycles. To make sure of the availability of current national data, the SDR is conducted and coordinated with the other SESTAT survey, NSCG. The degradation of any single component would jeopardize the integrity and value of the entire SESTAT system and integrated database.

Because the SDR is a longitudinal survey, conducting the survey less frequently would make it more difficult to locate the persons in the sample because of the mobility of the U.S. population. This would result in both a higher attrition rate as well as less reliable estimates. Also, NSF's biennial reports and 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 survey cycles 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 Congressionally mandated biennial reports prepared by the NSF.

7. Special Circumstances

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

8. Federal Register Announcement and Consultations Outside the Agency

Federal Register Announcement

The Federal Register Notice for the SDR appeared on June 1, 2012 (See Attachment 2). NSF received no public comment in response to the announcement by the closing date of July 31, 2012.

Consultations Outside the Agency

The National Center for Science and Engineering Statistics (NCSES) 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 first decade of 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. In order to accomplish this goal, NCSES 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 activities covered several areas: sampling frame, population coverage, sample design, survey content, data system design, data dissemination, and informed redesign of the SESTAT surveys. 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 on Redesign

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 since 2010.

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 Committee on National Statistics (CNSTAT) recommendation from a 2008 report on *Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics System*. To obtain feedback from the S&E community, NCSES conducted extensive outreach efforts with a broad audience to include 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 extensive outreach efforts with the S&E community, NCSES decided to discontinue the National Survey of Recent College Graduates (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 National Survey of College Graduates (NSCG), through the use of the American Community Survey, 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.
- In addition, NCSES held several discussions regarding plans for accelerating SDR data collection and data processing in an effort to release more timely data. Plans for maintaining the SDR's high response rate while shortening the data collection period include emphasizing participation via the Web survey, contacting sample members more frequently, and offering incentives earlier (if necessary). Plans to reduce the length of time required for data processing include starting data editing earlier and finalizing coding sooner.

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 2010 and 2013 survey rounds:

- NCSES has convened a Human Resources Experts Panel (HREP) to help improve data collection on the 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 accomplishes 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 NCSES 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 has met 7 times since it 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 with a selection of current and potential future licensees who presented their research findings and ideas to NSF as well as to the broader research community.
- The SESTAT surveys, particularly the SDR, 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 research, 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 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

No incentives will be offered to respondents in the initial stages of data collection. Based on current plans, however, in the later stages of data collection NSF will offer incentives to increase response rates across analysis groups, following the incentive plan that worked effectively in the 2008 and 2010 rounds. In order for estimates from the 2013 SDR to be available for publication in the 2015 *Women, Minorities, and Persons with Disabilities in Science and Engineering* report, adequate data must be obtained from all sampling strata in a timely manner. Because late-stage incentives have been shown to be effective in past SDRs to increase response rates among critical analytic domains, their use is planned for the 2013 SDR, as well. (NSF reserves the right to forego implementing incentives if it deems them unnecessary to produce appropriate estimates.)

The 2006 SDR included an experiment that demonstrated that sample members receiving a prepaid monetary incentive (\$25 check) had higher completion, cooperation, and response rates than did those not receiving an incentive. Moreover, data quality was not negatively impacted by the use of incentives. Based on these results, the 2008 SDR offered a late-stage incentive (\$30 prepaid) to 100% of nonrespondents in low responding sampling strata and to 20% of nonrespondents in higher responding strata to ensure sufficient data to produce survey estimates for critical analytic domains (as the sampling strata are aligned with the analysis domains). (See table 1.1 for 2008 SDR results.)

Table 1.1. Survey Results from Offering Late-Stage Incentives: 2008 SDR

Completed Survey	Cashed Check	Cases	Percent
Yes	No	884	21.0%
	Yes	2,057	48.8%
No	No	1,223	29.0%
	Yes	51	1.2%
Overall		4,215	100.0%

To increase the response rates of low responding sampling strata, achieve an overall response rate of 80%, and complete data collection on time, the 2010 SDR implemented an incentive protocol similar to the one used successfully in the 2008 SDR (and proposed again below for the 2013 SDR). Specifically, the 2010 SDR offered a late-stage incentive (\$30 prepaid) after 6.5 months of data collection. Nonrespondents in low responding sampling strata were offered the incentive with certainty (100%), and 20% of the nonrespondents in better performing strata were offered the incentive.

Of the 11,352 sample members eligible for a late-stage incentive, 10,412 were selected for the incentive. However, only 4,978 of these cases were not in locating (i.e., had current contacting information). In the end 7,174 sample members were offered the late-stage incentive and 2,383 (33%) cashed the check. The survey response and cashing outcomes for the 2010 SDR late-stage sample are displayed in Table 1.2. As shown, the completion rate for the incentive group was 55.4% (23.4% + 32.0%) compared to 44.7% for the non-incentive group, for a gain of 11%. More specifically, among the strata whose members received the incentive with certainty, the average response rate increased 14.5% (from 62.5% to 77.1%), whereas among the strata that had 20% of the nonrespondents incentivized, the average response rate increased 6.6% (from 78.1% to 84.7%).

Table 1.2. Survey Results for Located, Late-Stage Incentive Eligible Cases: 2010 SDR

Late-Stage Incentive Group	Survey Status	Check Status	Cases	Percent
No Incentive Offer	Nonrespondent	NA	350	55.3%
	Complete	NA	283	44.7%
Incentive Offer	Nonrespondent	Not cashed	3,113	43.4%
		Cashed	90	1.3%
	Complete	Not cashed	1,678	23.4%
		Cashed	2,293	32.0%

Proposed Plan for the 2013 SDR

The overall strategy for the late-stage incentive is to ensure that all sample members who have been subject to the standard survey data collection protocols and still remain as survey nonrespondents will have a probability of receiving a monetary incentive. In the plan used for the 2008 and 2010 SDR and again proposed here, a greater probability of selection for the incentive will be given to cases in those sampling cells where there are relatively lower response rates, in order to elicit sufficient data to produce survey estimates for critical analytic domains (given that the sampling cells are aligned with the domains of interest for analysis).

The sample parameters for the SDR are summarized below. There are 150 sampling cells, with three major stratification variables:

1. Demographic group (9 values) – this variable describes citizenship, disability status, and race/ethnicity,
2. Sex (2 values), and
3. Field of degree (7 values) – this variable describes the field of doctorate.

To allocate its available resources for the monetary incentive to late-stage survey nonrespondents, NSF will divide the SDR's sampling cells into two groups. One group will contain sampling cells from which all pending nonresponse cases will be offered an incentive, referred to as the "certainty" incentive treatment. The second group will be comprised of sampling cells from which a random sample of the pending nonresponse cases will be selected for the incentive offer, referred to as the "sample" incentive treatment. Whereas 100% of the pending nonresponse cases will be offered an incentive in the certainty incentive treatment, 20% of the cases in the sample incentive treatment will receive the same offer.

To define which sampling cells would fit into the certainty and sample incentive treatments, NSF will review the 2010 final response rates, the 2013 preliminary response rates, and the difference between the 2013 preliminary response rates and the target rate at the current stage in the 2013 data collection. Based on these data points, NSF will use the following three broad parameters to determine the incentive treatment.

1. Parameter 1: High vs. Low Response Rate in 2010
A "low" response rate is any rate that is 105% or less of the overall 2010 final weighted response rate (i.e., 80.5%). A "high" response rate is any rate that is higher than 105% of the overall 2010 response rate.
2. Parameter 2: Fast vs. Slow Response Rate in 2013
A "slow" response rate is any response rate that is 110% or less of the overall 2013 preliminary response rate. A "fast" response rate is any rate higher than 110% of the overall 2013 preliminary response rate.
3. Parameter 3: Big vs. Small Response Rate Difference between the Target Rate and the 2013 Preliminary Rates
A "big" difference is any response rate difference that is 10% lower than the overall difference between the target rate and the 2013 preliminary response rate. A "small" difference is any response rate difference that is less than 10% lower than the overall difference between the target rate and 2013 preliminary response rate.

To stay within the budgetary restraints of the 2013 SDR, given the response rates at the time when the incentives will be offered, it may be necessary to adjust the cut points for the three parameters. However,

the three parameters will remain the same as will the 20% of the sample incentive treatment group receiving the incentive.

Taking these three parameters together, eight scenarios are possible. Table 1.3 provides the composition of possible incentive treatment groups and the rationale for the incentive treatment for each scenario. Each 2013 SDR sampling cell will be assigned a “certainty” or “sample” incentive treatment status based on this plan.

Table 1.3. Response Rate Parameters and Assignment of Incentive Treatment

2010: High or Low Response Rate	2013: Fast or Slow Response Rate	Target-2013: Big or Small Difference	Group Acronym	Late Stage Incentive Treatment	Rationale for Incentive Treatment
Low	Slow	Small	LSS	Certainty	Cells that did not perform well in 2010, and are still not performing well relative to other cells in 2013.
Low	Slow	Big	LSB	Certainty	Cells that did not perform well in 2010, and are still not performing well relative to other cells in 2013.
Low	Fast	Small	LFS	Sample	Cells that did not perform well in 2010, but are performing well relative to other cells in 2013.
Low	Fast	Big	LFB	Certainty	Cells that did not perform well in 2010, and while they are performing better in 2013, still need to overcome a large difference to perform as well as the target rate.
High	Slow	Small	HSS	Sample	Cells that performed well in 2010, and although they are slow-performing in 2013, the difference to overcome is small in order to perform as well as the target rate.
High	Slow	Big	HSB	Certainty	Cells that performed well in 2010, but are not performing well in 2013, and the difference to overcome is large in order to perform as well as the target rate.
High	Fast	Small	HFS	Sample	Cells that performed well in 2010 and continue to do so in 2013.
High	Fast	Big	HFB	Sample	Cells that performed well in 2010 and 2013, even though the difference rate is large relative to the target.

Incentive Costs

If this incentive plan is implemented, NSF would offer a \$30 prepaid incentive to selected cases for the 2013 NSDR, as was done for the 2008 and 2010 NSDR. The total cost of incentives paid for the 2010 SDR was \$74,460.

Future Plans

The SDR incentive plan focuses on obtaining sufficient responses to reliably publish key analytical domains. Over the current cycle we intend to analyze the 2010 and 2013 SDR data to:

1. Evaluate the bias of key analytical domains
2. Assess the effect of incentives on bias
3. Adjust the SDR sample design for the 2015 cycle to accommodate flow processing in a responsive design setting so that incentives can be used more effectively for bias reduction.

10. Assurance of Confidentiality

The NSF and its contractors are fully committed to protecting the confidentiality of all survey respondents. SDR data will be collected under the authority of the National Science Foundation Act of 1950, as amended, and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002. Cover letters and survey questionnaires to each selected respondent advise them that the information they provide is confidential (see Attachment 3 – Proposed 2013 SDR Mailing Materials and Attachment 4 – Proposed 2013 SDR Questionnaire). The same notice of confidentiality will be used in the introduction to the CATI interview and will be displayed prior to the start of the survey in the Web instrument.

Standard data collection procedures incorporate numerous safeguards for the data and must conform to a detailed security plan approved by NSF. While collecting SDR data, the contractor separates information that could identify a particular sample member from data about that person. Each sample member is assigned a unique identifier, and this identifier is used to store identifying information (such as name, address, etc.) in a separate and secured database apart from the survey response database. The SDR affiliated contractors and NSF staff also receive annual CIPSEA training to reinforce their legal obligations to protect the privacy and confidentiality of the SDR data and staff must sign data use agreements annually to acknowledge this legal obligation.

SDR hard copy questionnaires and other contact materials are housed in a secured storage room at the contractor's production facility. Hard copy materials are accessed from the file room only by authorized staff and only when necessary for data collection activities. The contractor's electronic systems are on a local area network (LAN). All contractor systems used to store electronic survey data are secure by design and protected by passwords only available to authorized study staff.

The contractor takes special steps to ensure that data collected via the Web questionnaire are secure. First, access to the Web instrument is only allowed with a valid Personal Identification Number (PIN) and password correctly entered in combination. Second, data are transmitted by the Secure Sockets Layer (SSL) protocol that uses powerful encryption during transmission through the Internet. If a respondent keeps a Web survey open without any activity, the Web server closes it after a short period of inactivity, thus preserving the data up to the break-off point and securely closing the connection. The Web system architecture process is designed in a way that places authentication information and response data on physically separate servers. This strategy provides an extra layer of security to protect response data. Both development and production servers are backed up nightly, as the contractor's disaster recovery plan requires.

All data and analysis are reported in aggregate form only and measures are taken so that the identity of individuals or organizations is not disclosed.

11. Justification for Sensitive Questions

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

12. Estimate of Respondent Burden

A statistical sample of approximately 47,000 persons, identified as having a doctorate in a science,

engineering, or health field from a U.S. university will be selected for the 2013 SDR. This sample will include approximately 40,000 individuals residing in the U.S. (national component) and 7,000 residing abroad (international component). The amount of time to complete the questionnaire may vary depending on an individual's circumstances; however, on average it will take approximately 25 minutes to complete the survey. Assuming an 80% response rate (37,600 respondents) NSF estimates that the total burden for the 2013 SDR will be 15,667 hours.

The total cost to respondents for the 15,667 burden hours is estimated to be \$685,431. This is based on an estimated median annual salary of \$91,000 per full-time employed SDR respondent from the 2010 SDR data. Assuming a 40-hour workweek over 52-weeks of employment, this annual salary corresponds to an hourly rate of \$43.75.

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 the Federal Government

The total estimated cost to the Government for the 2013 SDR is \$6.6 million for survey cycle costs and for NSF staff costs to provide oversight and coordination with the other SESTAT survey. The cost estimate for the survey cycle is \$6.0 million, which is based on sample size; length of questionnaire; CATI and Web data collection technology; administrative, overhead, design, printing, mail and telephone data collection costs; 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; preparation of documentation and final reports; analysis, and tabulations. The NSF staff costs are estimated at \$562,500 (\$150,000 annual salary of 1.5 FTE for 2.5 years of the 2013 SDR survey cycle).

15. Reason for Change in Burden

The 2013 SDR will include a slightly larger sample size (from 45,700 in 2010 to 47,000 in 2013) to accommodate the additional subsample added to the ISDR panel. The change in burden hours from the 2010 SDR reflects the increase in the total SDR sample size.

16. Schedule for Information Collection and Publication

The NSF does not plan to use any complex analytical techniques in NSF publications using these 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 (Mail, CATI, Web)	February 2013 – July 2013
Coding and Data Editing	March 2013 – September 2013
Final Edited/Weighted/Imputed data file	December 2013
SDR Info Brief	Spring 2014
SDR Detailed Statistical Tables	Spring 2014
SDR Public Use File	Summer 2014

17. Display of OMB Expiration Date

The OMB Expiration Date will be displayed on the 2013 SDR questionnaire; in the Web survey version, it will be included on the informed consent page of the Web survey and available in a help

screen accessible at any point in the Web survey; in the telephone interview, it will be read to sample members during the introductory informed consent.

18. Exception to the Certification Statement

Not Applicable.