B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

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

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

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

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

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

2. STATISTICAL PROCEDURES

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

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

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

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

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

3. METHODS TO MAXIMIZE RESPONSE

Maximizing Response Rates

In order to maximize the overall survey response rate, NSF and the Census Bureau will implement procedures such as conducting extensive locating efforts and follow-up telephone interviews for nonrespondents to the mail questionnaire. To increase the response and minimize potential bias, a targeted monetary incentive will be offered to convert refusals from respondent groups that traditionally have a low response rate to the NSCG. A monetary incentive will also be included in an experiment to determine the conditioning effect in offering of incentives in subsequent survey responses. Once the details of this experiment are finalized, NSF will submit the experiment proposal for OMB approval.

The contact information obtained from the 2006 NSCG and 2006 NSRCG for the sample members and for the people who are likely to know the whereabouts of the sample members will be used to locate the sample members in 2008.

The Census Bureau will use a combination of locating and follow-up methods similar to the procedures used for the 2006 NSCG to maximize the survey response rate. The Census Bureau will utilize all of the available locating tools and resources to make the first contact with the sample person. The Census Bureau will use the U.S. Postal Service (USPS)'s automated National Change of Address (NCOA) database to update addresses for the sample. The NCOA incorporates all change of name/address orders submitted to the USPS nationwide, which is updated at least biweekly.

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

Dealing with Issues of Nonresponse Bias

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

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

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

In 2008, further assessment will be made of the extent of remaining bias by comparing weighted estimates for the survey sample that can be observed in the sample frame (e.g. degree field, degree level, and gender) to estimates for the population that the weighted sample is intended to represent.

4. TESTING OF PROCEDURES

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

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

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

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

Survey Questionnaire Review and Research

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

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

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

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

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

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

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

Pretest Phase I – Cognitive interviews

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

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

Pretest Phase II – Mail Field Test

The field test consisted of two mailings of NSCG and the other two SESTAT surveys with a reminder postcard in between; no further nonresponse follow-up was conducted due to time constraints. The NSCG mail pretest included a sample of 1,500 selected from a commercial list of 5,000 names of bachelor's degree holders with address, sex, age, and occupation information, and between the ages of 21 and 75. To mimic the proportion of science and engineering cases from the 1995 NSCG, MPR selected 15 percent of the cases from computer occupations, 20 percent from engineering occupations, and 65 percent from other occupations for a total of 1,500 sample members. Each sample member was randomly assigned to one of four control or experimental groups.

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

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

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

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

Mail Pretest Design

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

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

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

Survey Contact Materials

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

Questionnaire Layout

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

2006 Survey Methodology Tests

Postpaid Incentive Experiment

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

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

Reminder Experiment

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

collection phase because each reminder had an immediate effect in boosting the survey responses when administered.

Due Date Contact Experiment

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

Survey Methodology Tests to be Undertaken

As described in Section A, to better understand the effect of incentive conditioning on the survey panel, a monetary incentive experiment is being considered for the sample members who received a monetary incentive in the 2006 survey round. In addition to an incentive conditioning experiment, the NSF is considering offering a monetary incentive to the final refusals near the end of the data collection to minimize potential nonresponse bias.

Details on the incentive conditioning experiment plan are currently under development. The 2006 NSRCG panel sample members who received prepaid incentives with the first questionnaire mailing in 2006 will be split into treatment and control groups where only the treatment group will again receive an incentive with the first questionnaire mailing. The incentive plan will be designed to determine if the previous incentive receipt has any negative effect on the subsequent survey participation when no incentive is offered.

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

5. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION

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