

3145-0141 Revised Section B, National Survey of College Graduates

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

In the past, NSF has drawn a completely new sample for each decade from the decennial census long form. Beginning with the 2010 Census, the ACS replaced the long form. The NSF will use the ACS as a sampling frame for the 2010 NSCG and beyond. After reviewing numerous sample design options proposed by the NSF, the Committee on National Statistics (CNSTAT) recommended a rotating panel design for the 2010 decade of the NSCG (National Research Council, 2008). The use of the ACS as a sampling frame will allow the NSF to more efficiently target the S&E workforce population. Furthermore, the rotating panel design planned for the 2010 decade allows the NSCG to address certain deficiencies of the previous design including the undercoverage of key groups of interest such as foreign-degreed immigrants with S&E degrees.

The design for the 2010 decade sample will select more cases in small cells of particular interest to analysts, including underrepresented minorities, women, persons with disabilities and non-U.S. citizens. This will result in the surveys of the 2010-decade continuing to oversample underrepresented minorities, women, and persons with disabilities as in the 2000 decade design. The goal of this oversampling effort is to provide adequate sample for NSF's congressionally mandated report on Women, Minorities, and Persons with Disabilities in Science and Engineering.

To transition into the rotating panel design, the 2010 NSCG sample will include a subsample of the 2008 NSRCG and 2008 NSCG, and a sample from the ACS. The majority of 2010 NSCG sample, 65,000 cases, will be selected from the 2009 ACS. This group is referred to as the new cohort sample. The other portion of the 2010 NSCG sample, 35,000 cases, will be subsampled from the 2008 NSCG and 2008 National Survey of Recent College Graduates (NSRCG). This group is referred to as the old cohort sample.

The 2010 NSCG survey target population will consist of all U.S. residents under age 76 with at least a bachelor's degree as of January 1, 2009. The new cohort sample will provide complete coverage of this target population. The old cohort sample, on the other hand, will provide only partial coverage of the target population. Specifically, the old cohort will cover the population of U.S. residents with a bachelor's or master's degree in a SEH field earned in the U.S. as of June 30, 2007 and U.S. residents with any foreign-earned SEH degree as of April 1, 2000. Unbiased population estimates will be possible with careful weighting.

There are several advantages of this sample design. It will: 1) permit longitudinal analysis of the retained cases from the 2000-decade sample; 2) permit benchmarking of estimates to population totals derived from the sample using ACS; and 3) maintain the sample sizes of small populations of scientists and engineers of great interest such as underrepresented minorities, persons with disabilities and non U.S. citizens. .

There are two different versions of the questionnaire, one for each cohort. The main difference in the two questionnaires is the degree history questionnaire grid and certain demographic questions (e.g., race, ethnicity, and gender) that are not asked in the old cohort questionnaire, since this information has been collected for all old cohort sample cases in past survey cycles.

The target response rate for the new cohort is approximately 80 percent. The target response rate for the old cohort is approximately 90 percent. NSF estimated a response rate of 80% for the new cohort based on the following factors: 1) The contact information from the 2009 ACS based sample for the 2010 NSCG is much more recent than information from the 2000 Decennial long form respondents who were followed three years later in the 2003 NSCG. A significant proportion of the nonrespondents to NSCG in the past have been the un-locatable cases; 2) Availability of the web survey option will be more appealing to the younger college degreed population in the NSCG, which should help in increasing response rates from a demographic group that in past NSCG's had had lower response rates; and 3) Employment of mail/web mode initial survey contact, followed-up by additional reminder mailings mixed with CATI in 2010 is expected to raise response rates. The 2003 NSCG data collection was done sequentially where the initial mode was a mailed paper questionnaire, followed for mail nonrespondents by CATI and finally followed by CAPI for a small subsample of CATI nonrespondents. The CATI follow-up level of effort was somewhat limited in 2003 due to employment of the CAPI mode because Census Regional offices did not want too many CATI attempts before the cases were sent to them for CAPI.

2. SURVEY METHODOLOGY

The old cohort portion of the 2010 NSCG frame will be sampled separately from the new cohort portion. However, both sample designs will use similar stratification variables to the extent possible. Both cohorts will be stratified by highest degree level, a composite occupation variable (which is composed of S&E bachelor's field of degree and occupation), and a composite demographic variable (which is composed of race, ethnicity, disability status, citizenship, and foreign earned degree status). The multiway cross classification of these four stratification variables produces 567 possible sampling cells. This design ensures that the cells needed to produce the small demographic/degree field groups 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) will be maintained. Research on the panel design of NSCG sample allocation has shown that NSCG can produce the estimates for the key domain statistics that can support the reliability target in Table 1-5 (Attachment C). Therefore, 2010 NSCG reliability targets should be close to these targets for the S&E population. Final sample allocation for strata in both the old and new cohorts will be determined after analysis of the 2009 ACS sampling frame when it becomes available. The sample allocation will be determined based on reliability requirements for key NSCG analytical domains provided by the NSF.

Estimates from the 2010 NSCG will be based on standard weighting procedures. As was the case with sample selection, the weighting adjustments will occur separately for samples from each cohort. The goal of the separate weighting processing is to produce individual cohort final

weights that reflect the respective cohort population. To produce the individual cohort final weights, each case will start with a base weight defined as the probability of selection into the 2010 NSCG sample. This base weight reflects the differential sampling across strata. Base weights will be then be adjusted to account for noninterviews. After the noninterview adjustment, weights will be raked to ACS population totals through an iterative raking procedure that ensures population totals are upheld.

After the completion of the weighting, some of the weights may be relatively large compared to other weights in the same analytical domain. Since extreme weights can greatly increase the variance of survey estimates, NSF will examine weight trimming options. When weight trimming is used, the final survey estimates may be biased. However, by trimming the extreme weights, the assumption is that the decrease in variance will offset the associated increase in bias so that the final survey estimates have a smaller mean square error. Depending on the weighting truncation adjustment used to address extreme weights, it is possible the weighted totals for the key marginals will no longer equal the population totals used in the iterative raking procedure. To correct this possible inequality, the last step in the 2010 NSCG individual cohort weighting processing will be an additional execution of the iterative raking procedure. After the additional execution of the iterative raking procedure, the resulting weight will be the individual cohort final weight.

To increase the reliability of estimates of the small demographic/degree field groups used in the congressionally mandated report on Women, Minorities and Persons with Disabilities in Science and Engineering (See 42. U.S.C., 1885d), NSF will combine the new and old cohort together and form combined cohort weights. The combined cohort weights will be formed by adjusting the two sets of individual cohort final weights to account for the overlap in target population coverage. The result will be a combined cohort final weight for all 100,000 NSCG sample cases.

Replicate Weights. Two replicate weights for variance will also be constructed separately to allow for the old cohort sample and the new cohort sample. The variance for the combined cohort estimates will be constructed from these replicated weights. The entire weighting process applied to the full sample will be applied separately to each of the replicates in producing the replicate weights.

Standard Errors. The replication weights will be used to estimate the standard errors of the 2010 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 replicated 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 collect the survey data using three collection modes (mail, web, and CATI). The contact information obtained from the 2008 NSCG, 2008 NSRCG, and the 2009 ACS 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 2010.

The Census Bureau will refine and use a combination of locating and contact methods based on the past surveys 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 initial mailing to the NSCG sample members will include a paper questionnaire along with an option to complete the survey by web.

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.

A multimode data collection protocol will be used to improve the likelihood of gaining cooperation from sample cases that are located. Sample cases will be initially offered a choice to respond by paper or web questionnaire. Offering more than one response option communicates flexibility and consideration for the respondent, which may help obtain an increased number of responses. Nonrespondents will be followed in CATI. The college graduate population is mostly web-literate, so offering a web response option is apt to be appealing to NSCG respondents, especially the NSRCG panel sample members.

In addition to these procedures, the following steps will be taken to maximize response rates and minimize nonresponse:

- Developing "user friendly" survey materials that are simple to understand and use
- Sending attractive, personalized material using priority mail, making a reasonable request of the respondent's time, and making it easy for the respondent to comply
- Using priority mail for targeted mailings to improve the chances of reaching respondents and convincing them that the survey is important
- Devoting significant time to interviewer training on how to deal with problems related to nonresponse and ensuring that interviewers are appropriately supervised and monitored

- Using refusal-conversion strategies that specifically address the reason why a potential respondent has initially refused, and then training conversion specialists in effective counterarguments

See Appendices E and F for survey mailing materials.

4. TESTING OF PROCEDURES

For 2010, the new NSCG web instrument was tested and finalized based on the findings from the two rounds of usability testing conducted at the Census Bureau cognitive lab by the Statistical Research Division.

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.

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.

All content items in the SESTAT survey questionnaires had undergone an extensive review and testing before they were included in the 2008 questionnaire. The 2010 NSCG questionnaires will include no new items. They will include several module items rotating in from prior survey rounds and one new category in the core items on disability, which was taken from the ACS, although not previously fielded in the SESTAT surveys.

For 2010, the NSCG questionnaire content has been revised from 2008 as follows:

- Changed survey reference date from October 1, 2008 to October 1, 2010.
- Rotated in questions determining if and when a respondent was previously retired (last asked in 2003).
- Rotated in a question determining if the respondent's employer is a new business (last asked in 2003).
- Rotated in a module on respondent's job satisfaction with various job attributes (last asked in 2003).
- Rotated in a module on employer-provided job benefits (last asked in 1997).

- Rotated in a question on sources of Federal support with a reduced number of Federal agencies in the list (last asked in 2003).
- Rotated in questions about professional meeting attendance and association membership (last asked in 2003).
- Rotated in a module on respondent's ranking of the importance of various job attributes (last asked in 2003).
- Rotated in a module on reference week enrollment at college or university (last asked in 2003).
- Rotated in immigrant module questions (receipt year of permanent U.S. resident visa, year came to U.S., type of entry visa, reasons for coming to U.S., dual citizenship status) (last asked in 2003).
- Added a new category on the respondent's functional limitations with regards to concentrating, remembering or making decisions (from ACS questionnaire).
- For the new cohort, included core education history items (high school, community college, all degrees earned at a bachelor's or higher), and background demographic questions (place of birth, Hispanic type, race, parents' education) (as in 2003 baseline survey).
- Rotated out a module on second job (status, job description, job category, relatedness of second job to highest degree).
- Removed a category 'Chronic illness or disability' as a reason for working fewer than 35 hours per week due to too few cases reporting that reason in past survey cycles.
- Modified the format of telephone numbers of the respondents to specify home, work, or cell number for the daytime, evening and other telephone numbers.
- Added four new computer occupation codes and dropped one old code from the Job Category List based on the updated 2010 Standard Occupation Classifications (SOC).
- Modified the Field of Study List based on the updated 2010 Classification of Instructional Programs (CIP).

A complete list of questions proposed to be added, dropped, or modified in the 2010 NSCG questionnaires is included in Appendix D.

2008 Survey Methodology Test

Mode Preference Study

In an effort to boost response rates and reduce costs, a study was conducted in the 2008 NSCG to investigate respondents' mode preferences and the propensity to respond in the preferred mode.

In the 2006 NSCG, cases that responded in mail and Computer Assisted Telephone Interviewing (CATI) were asked in what mode they would prefer to answer subsequent surveys. In the 2006 National Survey of Recent College Graduates (NSRCG), the mode preference question was only asked of cases responding in CATI. For the 2008 NSCG mode preference study, the focus was placed on the cases that reported a telephone interview as their preferred data collection mode.

Through this experiment, it was concluded that offering the preferred mode seemed to have an initial impact on response rates, but did not have an impact in the overall response rates once all response mode options were taken into account. As seen with the initially higher response rate, by contacting the respondent in their preferred mode, a cost savings appeared to exist, as further follow-up was not needed. The 2010 NSCG will continue to target the respondent's preferred response mode.

2010 NSCG Web Survey Instrument Usability Testing

For the 2010 NSCG a web component is being introduced for the first time. The user interface is an important element of a Web-based instrument. For a questionnaire to be successful, the user interface must be able to meet the needs of its users in an efficient, effective, and satisfying way. Usability testing was conducted at the U.S. Census Bureau cognitive lab by the Statistical Research Division staff in the fall of 2009 and in early 2010 to examine the user interface of the NSCG web survey that will be accessible on the Internet. The testing evaluated the success and satisfaction of the test participants using an online version of the 2008 NSCG. The eye tracking software was used to test the long list collection, and the placement of navigational buttons. The test results indicated that the users are highly satisfied with the usability of the web instrument. The 2010 NSCG web component is being developed using the result of these studies.

Survey Methodology Tests to be Undertaken

Mode Study

Multi-mode surveys are increasingly becoming a norm in survey design as a means of maintaining response rates, improving coverage, and reducing costs. However, the data collection modes a survey uses can impact the survey results. Research has shown that people answer questions differently on a self-administered questionnaire (SAQ) than when interviewed over the phone or in person. An example of this phenomenon is the reporting of more positive responses to scale questions on telephone than on web-based SAQ surveys (Dillman & Christian, 2005; Christian, Dillman, & Smyth, 2008). If switching data collection modes produces different measurement, then the response rate and coverage gains may be offset by the undesirable changes in measurement (Dillman et al., 2008).

In recent years, the use of a web-based SAQ over the Internet has become an increasingly frequent mode of data collection. While research was conducted as part of the 1993 National Survey of College Graduates (NSCG) to examine the mode effect associated with the paper SAQ and telephone data collection modes, nothing is known about the potential effect a web-based SAQ might have on NSCG estimates. To investigate this issue, NSF will conduct a mode effect study as part of the 2010 NSCG production.

The main goal of this study will be to determine if the responses to the 2010 NSCG responses are affected by the mode of data collection. To reach this goal, the data obtained for the 2010 NSCG in the three modes, paper SAQ, CATI and web-based SAQ, will be compared. The following issues will be investigated:

- Do NSCG estimates for specific data items or types of items differ across data collection modes?
- Do unit nonresponse rates differ across data collection modes?
- Do item nonresponse rates differ across data collection modes?

This study will examine the 2010 NSCG data in an attempt to determine the impact collecting data in different data collection modes has on NSCG estimates, unit nonresponse rates, and item nonresponse rates. This evaluation is the first step in better understanding the impact of data collection mode on the NSCG estimates as we begin data collection in the 2010 decade. The hope is that the results from this analysis will enable us to better handle any mode effect in developing the 2012 data collection plans.

Additional Studies

NSF is considering a possible incentive research experiment in 2010. Any methodological tests we plan to conduct will be submitted for OMB approval prior to implementation.

5. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION

Chief consultant on statistical aspects of data collection is John M. Finamore (301) 763-5993, 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.