

## 3145-0020 Survey of Doctoral Recipients

### B. Collection Of Information Employing Statistical Methods

#### 1. Respondent Universe and Sampling Methods

The sampling frame/population for the 2010 SDR is defined by the Doctorate Records File (DRF). The DRF is a cumulative census of research doctorates awarded from U.S. institutions since 1920 which is annually updated through the Survey of Earned Doctorates (SED). The DRF contains approximately 1,746,000 individual-based records.

The sample design for 2010 will be consistent with the sample redesign developed and implemented in 2003 SDR and retained in both the 2006 and 2008 SDR that included drawing a subsample for the International Survey of Doctorate Recipients (ISDR). To be eligible for the 2010 National Survey of Doctorate Recipients (NSDR) target population, respondents have to:

- 1) have a doctoral degree in science, engineering or health from U.S. institutions awarded from academic years 1958 through 2009;
- 2) if a non-U.S. citizen, have indicated on the SED plans to stay in the U.S. after receiving their doctorate degree (or if a non-citizen from academic years 2001 through 2009, have indicated plans to leave the U.S. but were subsequently found to be living in the U.S.);
- 3) be non-institutionalized and under age 76;
- 4) be living in the U.S. as of October 1, 2010 (the survey reference date).

To be eligible for the 2010 ISDR target population, respondents have to:

- 1) a) have a doctoral degree in science, engineering or health from U.S. institutions awarded from academic years 1958 through 2009 if a U.S. citizen, b) from academic years 2001 through 2009 if a non-U.S. citizen, or c) from academic years 1958 through 2000 if a non-U.S. citizen who stayed in the U.S. until 2001;
- 2) if a non-U.S. citizen, have indicated on the SED plans to leave the U.S. after receiving their doctorate degree;
- 3) be non-institutionalized and under age 76;
- 4) be living outside the U.S. as of October 1, 2010 (the survey reference date).

For 2010, a sample will be selected from the new 2008-2009 doctoral cohort groups in the DRF and added to the longitudinal sample (which covers graduates through 2007) and is carried over from cycle to cycle. To offset this new cohort addition and to limit the overall sample size, a maintenance cut will be performed randomly on the NSDR longitudinal sample in addition to removing those who become ineligible due to advanced age, death or institutionalization and those who become eligible for the ISDR based on their residency abroad in 2008 (approximately 1,000 respondents). No cuts will be made on the ISDR longitudinal sample in 2010. Between the panel and the new cohort, the 2010 NSDR sample of 40,000 cases will be representative of approximately 815,000 individuals in the target population. Between the panel and new cohort, the 2010 ISDR sample of 4,700 will be representative of approximately 40,000 individuals in the target population. The targeted overall weighted response rate for the 2010 SDR is 80 percent. The plan for maximizing the response rate is presented in Section 3.

There are two types of SEH doctorate recipients who have been excluded from eligibility for the

SDR sample frame in the past but who may be eligible for the ISDR that began in 2003 : 1) non-U.S. citizens who reported plans in the SED to leave the U.S. after earning their SEH doctorate before academic year 2001 were considered ineligible for both the NSDR sample and the ISDR; 2) non-U.S. citizens who were selected into the NSDR but who had been found to reside outside of the U.S. for two or more survey cycles were also considered ineligible for the SDR prior to the 2003 SDR cycle but eligible for the international subsample as of the 2003 survey cycle if their degree was awarded in academic years 2001 or beyond (see Attachment 5 - Changes in the 2010 SDR methodology).

## 2. Statistical Procedures

As mentioned in the previous section, the 2010 sample design will be consistent with the 2003, 2006 and 2008 SDR sample designs. Stratification variables for the NSDR and ISDR samples include: demographic group, field of doctorate, and sex. The demographic group is a composite variable reflecting disability status, race/ethnicity, and citizenship-at-birth (U.S. or foreign).

The 2010 NSDR and ISDR samples will be selected across 194 sampling strata based on a multi-way cross of the stratification variables above. Sampling fractions for the NSDR and ISDR strata will differ between the existing cohort and the new cohort. The ISDR sample will include all existing cohort members (n=4,800) and a new cohort sample of 900 cases selected proportionately across ISDR strata from over 7,200 ISDR eligible cases from academic years 2008 and 2009. The ISDR/NSDR allocation is designed to minimize the design effects of respondents who change scope between NSDR and ISDR.

The NSDR and ISDR sample sizes and sampling design ensure the NSF will maintain the ability to produce the small demographic/degree field national 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). The 2010 ISDR sample is designed to ensure the NSF will produce reliable international estimates of U.S. trained doctorates working abroad by major geographic regions and broad fields of study.

Estimates from the 2010 NSDR/ISDR will be based on standard weighting procedures. As was the case with sample selection, the weighting adjustments will occur separately for cases for the old cohort and new cohorts. Each case will have a base weight defined as the probability of selection into the 2010 NSDR/ISDR sample. This base weight will reflect the differential sampling fractions across strata. For the NSDR old cohorts, the base weight will be equal to the final weight from the previous survey cycle, adjusted for the maintenance cut. The final analysis weights will be calculated in three stages:

- 1) First, a base weight will be calculated for every case in the sample to account for its selection probability under the sample design.
- 2) Second, an adjustment for unknown eligibility will be made to the base weight by distributing the weight of the unknown eligibility cases to the known eligibility cases proportionately to the observed eligibility rate within each adjustment class.
- 3) Third, an adjustment for nonresponse will be made to the adjusted base weight to account for the eligible sample cases for which no response was obtained.

*Replicate Weights.* A set of replicate weights based on the Balanced Repeated Replication (BRR) method will also be constructed to produce variance estimates. 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 BRR method will be used to estimate the standard errors of the 2010 SDR 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 out (replicated) several times. The dispersion of the resulting estimates then can be used to measure the variance of the full sample, including the national and international subpopulations.

### **3. Methods to Maximize Response**

#### ***Maximizing Response Rates***

The weighted response rate for the 2008 SDR was 80.5 percent. Extensive locating efforts, follow-up survey procedures and targeted data collection protocols will be used to maximize the survey response rate to maintain at least an approximately 80 percent response rate as a target for 2010. If necessary, NSF will propose offering a monetary incentive in the latter months of data collection (i.e., around the 5<sup>th</sup> month of the 9 month data collection period), and will submit an incentive plan to OMB for approval.

#### ***Locating***

The contact information obtained from the 2008 SDR and from the 2008 and 2009 SED surveys for the sample members as well as the people who are likely to know the whereabouts of the sample members will be used to locate the sample members in 2010.

The U.S. Postal Service's (USPS) automated National Change of Address (NCOA) database will be used 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. Vendors also maintain up to 36-months of historical records of previous address changes. It will also be used to track persons who have moved from their previous address at the time of 2008 survey. The names and addresses of mail nonrespondents will be matched to the most recent NCOA address updates with a vendor who appends telephone numbers.

The locating efforts will also utilize a specially-trained locating team who has proven themselves successful at searching for and finding nonrespondents with problem addresses or telephone numbers, especially those living abroad. Their locating strategy will include contacting employers, educational institutions, alumni associations, online publication searches, change of address searches, Directory Assistance and administrative record searches. In addition to last known address, locators have past contacting information available as far back as 2001. Locators will also have access to contact names and addresses given by respondents in past survey rounds, where available. An automated commercial telephone number matching service and the national death registry will also be used.

#### ***Data Collection Strategies***

NORC will continue to use a multi-mode data collection protocol (mail, CATI, web) to facilitate participation, data completeness and sample members' satisfaction.

A core set of contact materials (Prenotice Letter, Thank You/Reminder Postcard, and Cover Letters accompanying the SAQ) will be used in mailing to the SDR sample members (see Attachment 3 – Proposed 2010 SDR Mailing Materials). These contact materials will be tailored to address the particular issues or concerns of the sample groups to whom they are targeted. Tailoring will be based on type of cohort (2008 panel member versus new cohort), response in the past round, citizenship, retirement status and expressed mode preference. NORC will also utilize email versions of the contacting materials for sample members with email addresses on file.

NORC will conduct CATI follow-ups for those sample members who do not submit a completed questionnaire via a paper or Web form. The CATI Interviewing team will include Refusal Avoidance and Conversion specialists who have a proven ability to work with sample members to obtain their consent and participation.

### ***Incentive Plan for 2010***

In 2003, 2006 and 2008 SDR, several incentive experiments were conducted. In 2003, the experiments showed that a monetary incentive significantly improved response at the end of the field period compared to a non-monetary incentive or no incentive. Further, the response to a lower value prepaid incentive was better than a higher value postpaid incentive. In 2006 and 2008, a prepaid incentive of \$30 among late responders was effective in boosting the lagging response rate in the latter months of data collection. In 2010 no incentives will be offered to the international sample members. Should NSF decide an incentive is necessary in 2010 for the NSDR, a comprehensive incentive protocol will be submitted to OMB for approval in the 3<sup>rd</sup> or 4<sup>th</sup> month of data collection.

## **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.

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 before they were included in the final version of the 2008 questionnaires. The 2010 NSRCG questionnaires will include no new items. It will include several module items rotating in from prior survey rounds and one new category in the core item on disability, which was taken from the ACS although not previously fielded in the SESTAT surveys.

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

- Survey reference date changed from October 1, 2008 to October 1, 2010.
- Rotated out a module on publishing and presentation productivity and patents, which was asked in 2008.
- Rotated out a module on second job (status, job description, job category, relatedness of second job to highest degree), which was asked in 2008.
- Rotated in a module on college enrollment on the reference date, which was last asked in 2003.
- Rotated in a question about specific aspects of job satisfaction which was last asked in 2003.

A complete list of questions proposed to be added, dropped, or modified in the 2010 SDR questionnaire is included in Attachment 7.

### ***Survey Contact Materials***

Survey contacting materials will be tailored to best fit sample members need for information about the SDR and gain their cooperation. Materials requesting sample member participation via the Web survey will include access to the survey online. As has been done since the 2003 SDR, NSF and NORC will develop 2010 SDR letterhead stationery that includes project and NSF website information, and NORC's project toll-free telephone line, USPS and email addresses. Additionally, the stationery will contain a watermark that shows the survey's logo to help brand the communication for sample members for ease of recognition.

### ***Questionnaire Layout***

No change for 2010.

### ***Web-Based Survey Instrument***

Because of technological improvements and the wide proliferation of Internet users, offering a Web option to SDR respondents has become both feasible and desirable. The Web mode has become a valuable asset to the survey with regard to decreased cost and enhanced respondent satisfaction. In the 2003 SDR, this new mode was carefully introduced to avoid having a negative impact on the response rate or the high data quality that the SDR project has realized over the years.

The 2010 SDR will maintain the same functionality and software design as used in the 2003, 2006, and 2008 survey rounds. However, due to questionnaire changes, it will be necessary to recode some portions of the instrument. This development will take place during Spring 2010, and full testing of the reinstated questions as well as the entire instrument will be completed during August and September 2010. At least 60 percent of the SDR respondents are expected to participate by web based on their stated preference in the last round and the observed rate of web participation among the newly added cohort members in the last survey cycle.

## **5. Contacts for Statistical Aspects of Data Collection**

SRS Chief Statistician, Stephen Cohen, has overall responsibility of statistical aspects of the survey. Consultation on statistical aspects of sample design was provided by Brenda Cox, (703-875-2983, Senior Staff, Battelle) and Rachel Harter, (312-759-4025, Statistics and Methods Vice

President, NORC). At the NSF the contacts for statistical aspects of data collection are Lynn Milan (703 292-2275, SDR Project Officer) and Stephen Cohen (703-292-7769, SRS Chief Statistician).