B. Collection of Information Employing Statistical Methods

1. Respondent Universe and Sampling Methods

The sampling frame/population for the 2013 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 2010 DRF contains approximately 1,843,000 individual-based records.

The sample design for 2013 will be consistent with the sample design implemented in the 2010 SDR. The sample will cover all U.S. educated research doctorates in science, engineering, and health with two components: (1) those in the U.S., referred to as the National Survey of Doctorate Recipients (NSDR) and (2) those abroad, the International Survey of Doctorate Recipients (ISDR).

The sample design will be an integrated design that will minimize the design effect differences between NSDR and ISDR samples. This will minimize the impact on NSDR variances of misclassified sample members, e.g., those sampled in the ISDR who turn out to be residing in the U.S. In addition, U.S. citizens residing outside the U.S. will be included in the estimates of those residing abroad.

To be eligible for the 2013 NSDR sampling frame, sample members must:

1) have a research doctoral degree in science, engineering or health from a U.S. institution awarded from academic years 1958 through 2011;
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-U.S. 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 expected to be living in the U.S. as of February 1, 2013 (the survey reference date).

To be eligible for the 2013 ISDR sampling frame, sample members must:

1) a) have a research doctoral degree in science, engineering or health from a U.S. institution awarded from academic years 1958 through 2011 if a U.S. citizen, b) from academic years 2001 through 2011 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) for new cohort members, 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 expected to be living outside the U.S. as of February 1, 2013 (the survey reference date).

For 2013, a sample will be selected from the new doctoral cohort groups in the DRF (academic years 2010-2011) and added to the longitudinal panel sample, which covers graduates through 2009 and is carried over from cycle to cycle (see table 2). To offset this new cohort addition and to limit the overall sample size, a maintenance cut (estimated to be approximately 2,100 cases) will be performed randomly on the NSDR longitudinal sample, in addition to removing those who become ineligible due to advanced age, death or institutionalization (1,032) and those who become eligible for the ISDR based on their residency abroad in 2010 (693). (All nonrespondents from past rounds continue to be in the panel except for those who are selected into the maintenance cut or those who become ineligible, as explained above. However, sample members who complained to Congress about participation are not contacted in future rounds.) Approximately 3,800 new cohort cases will be

added to the 2013 NSDR sample to retain an overall sample size of 40,000. Between the panel and the new cohort, the 2013 NSDR sample of 40,000 cases will be representative of approximately 870,000 individuals in the target population.

No cuts will be made to the ISDR longitudinal sample in 2013 to continue building up its size. From the new cohort, 900 cases will be added, resulting in a total ISDR sample size of roughly 7,100 cases. Between the panel and new cohort, the 2013 ISDR sample of 7,100 will be representative of approximately 75,000 individuals in the target population. The targeted overall weighted response rate for the 2013 SDR is 80 percent. The plan for maximizing the response rate is presented in Section 3.

Table 2. 2013 SDR Sampling Frame by Cohort

Sample Frame Component	Old Cohort	New Cohort			Total	Final Sample Size
		Total New Cohort	2010 SED	2011 SED		
NSDR	38,424	63,955	31,300	32,655	102,379	40,000
ISDR	6,178	7,951	3,942	4,009	14,129	7,078*
Overall	44,602	71,906	35,242	36,664	116,508	47,078

*Estimate

2. Statistical Procedures

As mentioned in the previous section, the 2013 sample design will be consistent with the 2010 SDR sample design. 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 2013 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=5,700) and a new cohort sample of approximately 1,000 cases selected proportionately across ISDR strata. 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 2013 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 2013 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 2013 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 based on a propensity model.
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 2013 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 2010 SDR was 80 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 2013. If necessary, NSF will offer a monetary incentive in the latter months of data collection.

Locating

The contact information obtained from the 2010 SDR and from the 2010 and 2011 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 2013.

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 the 2010 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.

For locating ISDR members, the contractor will hire staff with foreign language skills (e.g., Mandarin, Cantonese, Korean, Russian, Arabic) to perform specialized tasks, such as navigating non-English websites, speaking with household members abroad, and searching Web databases hosted by government-funded science and engineering organizations in foreign countries.

Data Collection Strategies

The contractor will 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 2013 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 (2010 panel member versus new cohort), response in the past round, citizenship, retirement status and expressed mode preference. The contractor will also utilize email versions of the contacting materials for sample members with email addresses on file and attempt to increase participation via the Web.

The contractor 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 2013

In the 2003-2010 SDR survey round, 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 the 2006 SDR an incentive experiment demonstrated that incentives were effective at increasing completion rates without negatively impacting data quality. In 2008 and 2010, a prepaid incentive of $30 among late responders was effective in boosting the lagging response rate in the latter months of data collection. For the 2013 SDR, if incentives are deemed necessary by NSF to increase the response rates of critical domains, the protocol implemented in the 2008 and 2010 SDR will be followed (see section A9).

4. Testing of Procedures

Because data from both 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 two 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 the 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 2013 questionnaires. Compared to the 2010 SDR questionnaire, the 2013 version contains no new content. The 2013 NSCG questionnaire includes two new components: items about community college attendance and about funds borrowed/owed to finance one's education.

For 2013, the SDR questionnaire has been revised from 2010 as follows:

- Survey reference date changed from October 1, 2010 to February 1, 2013.
- Changed the response options for question E6.
- Modified the response options for questions A11 and A43.

A complete list of modifications to the 2013 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 the contractor will develop 2013 SDR letterhead stationery that includes project and NSF website information, and the contractor'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 2013.

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 2013 SDR will maintain the same functionality as in past rounds. At least 63 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 in the last survey cycle. (See Attachment 7 for survey items asked only in the Web (and CATI) instruments to capture information lacking in sample members' records.)

5. Contacts for Statistical Aspects of Data Collection

NCSES Chief Statistician, Stephen Cohen, has overall responsibility for statistical aspects of the survey. Consultation on statistical aspects of sample design was provided by Brenda Cox, (703-875-2983, Senior Staff, Battelle) and Michael Sinclair, (301-634-9493 Senior Fellow, Center for Excellence in Survey Research, 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, NCSES Chief Statistician).