

## B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

### 1. RESPONDENT UNIVERSE AND SAMPLING METHODS

The 2017 SDR features a slight sample size increase, from 120,000 cases in 2015 to 124,580 cases in 2017. The sample size increase reflects that along with the new cohort sample, the entire 2015 panel that remained eligible for the next cycle is included in the 2017 sample.

#### 1.1 Frame

The primary sampling frame source for the SDR is the Doctorate Record File (DRF). The DRF is a cumulative file of the count of research doctorates awarded from U.S. institutions since 1920. It is annually updated with new research doctorate recipients through the SED. The 2015 SDR sample selected from the 2013 DRF represented a surviving population of approximately 1.05 million individuals with SEH doctorates who were less than 76 years of age. The 2017 SDR is expected to represent a slightly larger and growing population of approximately 1.1 million SEH doctorates from the 2015 DRF including over 80,000 from the most recent two academic years. The 2015 DRF contains 2,104,870 records in total.

The target population for the 2017 SDR includes individuals who must:

- Have a research doctoral degree in a science, engineering or health field from a U.S. institution, awarded no later than academic year 2015, and
- Be less than 76 years of age on 1 February 2017 based on their date of birth.

The final 2017 SDR sampling frame consists of 196,336 cases classified into two groups as shown in Table 1.

1. Frame Group 1 contains the eligible cases from the 2015 SDR panel sample.
2. Frame Group 2 contains individuals who have earned their degree since the 2015 SDR sampling frame was built. These individuals have not previously had a chance to be selected for the SDR since they have just been awarded their doctorate degrees in the 2014 and 2015 academic years.

**Table 1: The 2017 SDR Frame Cases by Sample Component**

| Sample Component | Frame Group | Description  | SED Academic Years (AY) | Frame Cases | Population Size |
|------------------|-------------|--|-------------------------|-------------|-----------------|
| Panel            | 1           | 2015 SDR sampled cases that remain eligible for 2017 | 1960-2011               | 113,814     | 1,035,376       |
| New Cohort       | 2           | New cohort cases from SED AY 2014 and 2015           | 2014-2015               | 82,522      | 82,522          |
| Total            |             |  |                         | 196,336     | 1,117,898       |

#### 1.2 2017 Sample Design

The SDR historically featured a stratified systematic sample design, where the strata were defined by broad degree field, gender, race, ethnicity, citizenship, disability status, and other relevant demographic variables. As part of the 2015 SDR redesign, the sampling strata were defined as fine fields of degree (FFODs). The 2017 SDR will follow the 2015 SDR design, featuring a total of 223 strata. Again, the strata are defined by the SED FFOD alone with classification variables within strata, reflecting the emphasis on the new analytical objectives at the fine field level established for 2015 SDR. Each stratum corresponds to a fine field of degree, except the last stratum, which consists of a group of fine fields that no longer exist after the 2000 SED. These fields are grouped into one single stratum because they individually do not constitute analysis domains that contribute to the 2015

design analytical objectives. A total of 123,736 cases were allocated to the 222 strata that constitute the fine fields; while 844 cases remain in the 223rd stratum representing the discontinued fields after the 2000 SED.

Following the 2015 SDR design, the 2017 SDR also features oversampling of under-represented minorities (URM) and women within each sampling stratum. Oversampling of URM and women allows the 2017 sample to sustain the estimation capabilities under the 2013 and prior SDR design. The 2017 panel sample selected in the 2015 survey cycle also features oversampling of cases in earlier waves of the SDR to support limited but important longitudinal analysis. For the new cohort sample selection, demographic variables used for stratification in previous sample designs, such as citizenship at birth, predicted resident location, and doctorate award year, are used as sorting variables within each stratum to impose an implicit stratification to improve their representation in the sample.

The 2017 SDR sample allocation starts with a two-step process developed for the 2015 SDR sample to derive the stratum target allocation. The first step assigns an equal allocation to each fine field to meet a pre-specified level of precision at the fine field level; the second step allocates the remaining sample to bring the overall sample closer to a proportional representation of the 15 SDR broad field categories. The two-step allocation was implemented because broad fields with a large population but consisting of a small number of fine fields (e.g., Computer/Information Sciences) are underrepresented by the equal allocation in the first step. Similarly, broad fields with a small population but consisting of a large number of fine fields (e.g., Agricultural Sciences) are overrepresented. The two-step allocation makes the representation of broad degree fields more proportional to the population and minimizes the variation in sample weights for the full sample. The total target allocation to each fine field is then multiplied by the proportion of the new cohort population within the fine field to determine the new cohort sample allocation (See Attachment F for 2017 SDR Sample Allocation and Sample Selection Tables).

For 2015, the overall design effect was 2.12, reflecting a more disproportional sample allocation under the new design than under the 2013 design ( $d_{eff}=1.15$ ). Since the 2017 design is similar to the 2015 design, a similar design effect is anticipated. The fixed allocation to strata in the first step led to more weight variations because the frame size varies greatly across strata, while the second step allocation helped to control the design effects.

## **2. STATISTICAL PROCEDURES**

The SDR statistical data processing procedures have several components, including weighting adjustments to compensate for the complex design features, missing data imputation, and variance estimation.

### **2.1 Weighting**

A final weight will be computed for each completed interview to reduce potential bias in the estimates of population characteristics. Estimation bias could result from various sources, including unequal selection probabilities, nonresponse, and frame coverage issues. The weighting procedures will address all these factors through a series of adjustments to the sampling weights under the 2017 SDR design.

For a sample member  $j$ , its sampling weight will be computed as

$$w_j = \frac{1}{p_j}$$

where  $p_j$  is the inclusion probability under the sample design.

The sampling weight will be adjusted in sequence for unknown eligibility, unit nonresponse, and frame coverage based on similar methodologies developed for the 2015 SDR. First, for cases whose eligibility status is not determined by the end of the survey, their assigned base weights are transferred to cases whose eligibility is known. Next, among eligible cases, the weights of nonrespondents are transferred to the respondents so that the respondents represent all eligible cases in the sample. Finally, a raking adjustment aligns the sample to the frame population so that the sample estimates agree with the frame counts with respect to factors not explicitly controlled in the sample design.

Like the 2015 SDR, logistic regression models will be used to derive unknown eligibility and nonresponse weighting adjustment factors for different segments of the sample. Predicted propensity scores will be used to define weighting classes, and extreme weights will be trimmed to reduce the variation of the weights prior to raking. With a final weight, the Horvitz-Thompson estimator will be used to derive point estimates for various SDR variables.

## **2.2 Item Nonresponse Adjustment**

Historically, the SDR has conducted comprehensive imputation to fill in item-level missing data. Two general methods of imputation, logical imputation and hot deck imputation, have been used. The logical imputation method is usually employed during the data editing process when the answer to a missing item can be deduced from past data, or from other responses from the same respondent. For those items still missing after logical imputation, a hot deck imputation method is employed. In hot-deck imputation, data provided by a donor respondent in the current cycle is used to impute missing data for a respondent who is similar to the donor respondent based on a propensity model. The 2017 SDR will use similar imputation techniques, although the actual imputation models may differ in that longitudinal-based data from the 2015 cycle may be used to identify donors.

## **2.3 Variance Estimation**

The SDR has adopted a Successive Difference Replication Method (SDRM) for variance estimation. The SDRM method was designed to be used with systematic samples when the sort order of the sample is informative. This is the case for the 2017 SDR, which employs systematic sampling after sorting cases within each stratum by selected demographic variables. In 2015, a total of 104 replicates were used. Within each replicate, the final weight is developed using the same procedures applied to the full sample. The replicate weights also included a finite population correction factor to account for high sampling rates observed in smaller fine field strata. The results of SDRM are multiple sets of replicate weights that can be used to derive the variance of point estimates by using special variance estimation software packages such as SAS and SUDAAN.

The same variance estimation approach will be adapted for the 2017 SDR.

# **3. METHODS TO MAXIMIZE RESPONSE**

## **3.1 Maximizing Response Rates**

The weighted response rate for the 2015 SDR was 66.0% (unweighted, 67.9). Extensive locating efforts, nonresponse follow-up survey procedures, and targeted data collection protocols will be used to attain a targeted 75% response rate for 2017. An early incentive as outlined in section A.9 will be offered, in addition to a late-stage incentive in the latter months of data collection.

## **3.2 Locating**

Panel sample members who were categorized as locating problems in 2015 and new sample members with incomplete contacting data will first need to be located before a request for survey participation can be made. The

2017 SDR will follow a locating protocol similar to that implemented in 2015. The contacting information obtained from the 2015 SDR and prior cycles will be used to locate and contact the panel; the information from the SED will be the starting information used to locate and contact any new cohort cases in 2017.

**2017 SDR Locating Protocol Overview.** As in the prior SDR cycles, there will be two phases of locating for the 2017 SDR: prefield and main locating. Prefield locating activities include Accurint®<sup>14</sup> batch processing and individual searches, address review, and individual case locating (also called manual locating). Prefield locating occurs before the start of data collection and is used to ensure the initial outreach request for survey participation is sent to as many sample members as possible. Prefield individual case locating includes online searches, limited telephone calls to sample members, and telephone calls and emails to contact persons who may know how to reach the sample members. Main locating includes manual locating and additional Accurint® processing as needed. Main locating activity will begin at the start of data collection and will include contact (by mail, telephone, or email) with sample members and other contact persons. Both the prefield and main locating activities will be supported by an integrated online case management system. The case management system will include background information for each case, all the locating leads, all searches conducted, and all outreach attempts made which lead to the newly found contacting information (including mailing addresses, telephone numbers, and email addresses).

The 2017 SDR will implement an adaptive design methodology which will assess the locating rates and survey response by key analytic domains to tailor appropriate follow-up responses and late-stage incentive offers.

**Prefield Locating Activities.** The prefield locating activities consist of four major components, as described below.

1. For the panel (sample component 1) and the new cohort (sample component 2), 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 for residential addresses; this database is updated biweekly. The NCOA database maintains up to 36 months of historical records of previous address changes. However, the NCOA updates will be less effective for the new sample (sample component 2) since the starting contacting information from SED could be up to three years out of date.
2. After implementing the NCOA updates for the panel and new cohort, the sample will be assessed to determine which cases require prefield locating. This assessment is different for the panel cases than for the new cohort sample components.

Prefield locating will be conducted on panel cases which could not be found in the prior round of data collection or ended the round with unknown eligibility (meaning we did not successfully reach the sample member). An Accurint® batch search also will be run using the available information as necessary.

For the new cohort, an Accurint® batch search will be run using the available information provided in the SED. The returned results from Accurint® will be assessed to determine which cases are ready for contacting and which require prefield locating. There are four potential data return outcomes from the Accurint® batch search:

---

<sup>14</sup> Accurint® is a widely accepted locate-and-research tool available to government, law enforcement, and commercial customers. Address searches can be run in batch or individually, and the query does not leave a trace in the credit record of the sample person being located. In addition to updated address and telephone number information, Accurint® returns deceased status updates.

- a. Returned with a date of death. For those cases that return a date of death, the mortality status will be confirmed with an independent online source and finalized as deceased. When the deceased status cannot be confirmed, the cases will be queued for manual prefield locating and the possible deceased outcome will be noted in the case record so further searching on the possible date of death may be conducted.
  - b. Returned with existing address confirmed. For cases where Accurint® confirms the SED address as current (i.e., less than two years old), the case will be considered ready for data collection and will not receive prefield locating.
  - c. Returned with no new information. For cases where Accurint® provides no new information or the date associated with new contacting information is more than two years out of date, the cases will be queued for manual prefield locating.
  - d. Returned with new information. When Accurint® provides new and current contacting information, the new information will be used, the case will be considered ready for data collection, and will not receive prefield locating.
3. A specially trained locating team will conduct online searches and make limited calls to sample members and outreach contact persons as part of the manual locating effort throughout prefield locating, for those individuals not found via the automated searches. Only publicly available data will be accessed during the online searches. The locating staff will use search strategies that effectively combine and triangulate the sample member's earned degree and academic institution information, demographic information, prior address information, any return information from Accurint®, and information about any nominated contact persons. From the search results, locators will search employer directories, education institutions sites, alumni and professional association lists, white pages listings, real estate databases, online publication databases (including those with dissertations), online voting records, and other administrative sources. Locating staff will be carefully trained to verify they have found the correct sample member by using personal identifying information such as name and date of birth, academic history, and past address information from the SED and the SDR (where it exists).
  4. Additionally, the 2017 SDR will use Accurint® to conduct individual matched searches or AIM searches. AIM allows locators to search on partial combinations of identifying information to obtain an individual's full address history and discover critical name changes. This method has shown in other studies, as well as the 2015 SDR, to be a cost-effective strategy when locating respondents with out-of-date contact information. The AIM searching method will be implemented by the most expert locating staff and will be conducted on the subset of cases not found with regular online searches.

**Main Locating Activities.** Cases worked in main locating will include those not found during the prefield locating period as well as cases determined to have outdated or incorrect contacting information from failed data collection outreach activities. Prior to beginning the main locating work, locating staff who worked during the prefield period will receive a refresher training which focuses on maintaining sample member confidentiality particularly when making phone calls, supplementing online searches with direct outreach to the sample members and other individuals, and gaining the cooperation of those sample members and other individuals successfully reached. The locating staff will continue to use and expand upon the online searching methods from the prefield period and, ideally, gain survey cooperation from the found individuals. In addition to outreach to the sample members, main locating activities during data collection will include calls and emails to dissertation advisors, employers, alumni associations, and other individuals who may know how to reach the sample member.

### 3.3 Data Collection Strategies

A multi-mode data collection protocol (web, mail, and CATI) will be used to facilitate survey participation, data completeness, and sample member satisfaction. The 2017 SDR data collection protocols and contacting methods are built upon the methodology used in 2015. The data collection field period is segmented into four phases: a “starting” phase, “interim” phase, “late-stage” phase, and “last chance” phase. The starting and interim phases include four separate data collection protocols tailored to different sample groups. In the late-stage and last chance phases, all remaining nonresponse cases (regardless of their starting data collection protocol) receive a tailored contacting protocol based on assigned priority group.

The majority of the sample will be assigned to the web starting data collection protocol. However, some panel sample members will be assigned to the alternative modes based on their reported mode preferences, past response behaviors, and available contacting information. The four different starting protocols are implemented in tandem. The starting protocols and sample members assigned to these starting protocols are described below.

1. Web – This is the primary data collection mode and most cases start with the web protocol. The initial request to complete the 2017 SDR will be made by a USPS and/or email message that includes links to the online version of the survey; when both a USPS and email address are available, sample members are contacted by both means rather than one. This contacting strategy was tested in a 2013 methods experiment and works well. This starting protocol group includes the following sample members:
  - Cooperative respondents who prefer the web questionnaire,
  - Cooperative respondents who prefer the mail questionnaire, but have an email and mailing address,
  - New cohort sample members with complete sampling stratification variables who have a portable email address, (e.g., gmail or yahoo),
  - All locating problem cases as they are found, and
  - Cases with previously experienced language problems.
2. Mail – The initial request to complete the 2017 SDR will be made through a USPS mailing that includes a paper version of the survey. This starting protocol group includes 2015 panel sample members who reported they prefer the mail mode and do not have an email address, and all non-cooperative retirees. New cohort sample members without a portable email address will also receive the mail starting protocol.
3. Reluctant Mail – The initial request to complete the 2017 SDR will be made through a USPS mailing that includes a paper version of the survey. This protocol is a modified version of the starting mail protocol, but has fewer contacts with more time between contacts during the starting phase. This group will include panel sample members who are known to be reluctant survey participants – specifically, individuals who previously indicated they would complete the survey only after receiving an incentive, and panel sample members who refused to participate in 2015.
4. CATI – The initial request to complete the 2017 SDR will be made by a trained telephone interviewer who will attempt to complete the survey via CATI. This starting group includes 2015 panel sample members who reported they prefer the CATI mode, new cohort sample members with incomplete sampling stratification variables, institutionalized sample members, and other sample members whose only current contacting information is a valid telephone number at the beginning of data collection.

A core set of contact materials (Prenotice Letter, Thank You/Reminder Postcard, and Cover Letters accompanying the paper questionnaire) will be used as outreach to the SDR sample members (see Attachment E – Draft 2017 SDR Survey Mailing Materials). These contact materials are tailored to address the issues or concerns of the sample groups to whom they are targeted. Tailoring is primarily based on type of cohort (e.g., 2015 panel

or new cohort). Additional tailoring for the 2015 panel members is based on response/nonresponse in the past round, citizenship, retirement status, and expressed mode preference. Email versions of contact materials will be developed to communicate with sample members who have email addresses on file.

The type and timing of contacts for each starting data collection protocol is shown in Figure 1. The outreach method and schedule is consistent with the approach used in the 2015 cycle, but with some improvements described below.

- As in the 2015 cycle, the 2017 reluctant mail starting protocol will introduce the short or critical item only (CIO) version of the survey earlier in an effort to a) increase the survey participation rate for these cases, b) shorten their time to respond, c) decrease the overall number of contacting attempts, and d) reduce the need to offer an incentive in order to obtain the short version of the survey. Past experience demonstrates the SDR obtains a better unit survey response from reluctant sample members when the CIO survey is offered. Reluctant sample members have low item-nonresponse when completing the CIO survey.
- Of the cases eligible for 2017, 235 nonresponse panel members were classified as “hostile” refusals. These individuals were vehement and/or profane when declining to participate. These panel sample members will receive a single questionnaire mailing with a cover letter that acknowledges their refusal status and offers the web CIO. The 128 nonresponse panel members classified as “congressional” refusals will not be contacted in 2017. This is the protocol utilized in 2015 for both of these groups.
- As noted above, the new cohort sample component and newly found panel cases will be assigned to the web starting protocol. Research<sup>15</sup> shows that including a non-monetary token incentive in the initial request for survey participation improves survey response. To remain cost efficient while providing a token incentive, the initial USPS mailing with the request for survey participation to these sample members will include a four-color print information card explaining the purpose and value of the SDR. The information card will include a detachable panel that may be used as a bookmark. For confidentiality reasons the detachable bookmark will not mention the survey, but will be graphically appealing and will include useful information like measurement conversions.
- Monetary incentives will be offered during the starting data collection phase as described in section B3.4.

Telephone follow-up contacts will be conducted for those sample members who do not submit a completed questionnaire via a paper or online survey. To facilitate the multi-mode effort, the telephone call case management module will have the ability to record updated USPS and email addresses for sample members who request a paper survey or web survey access, respectively. The telephone interviewing team will include Refusal Avoidance and Conversion specialists who have a proven ability to work with doctoral sample members to obtain survey participation.

The overall 2017 SDR schedule of contacts by starting protocol is shown in Figure 1.

---

<sup>15</sup> Millar, M.M. and D.A. Dillman (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, 55 (2), 249-269.

**Figure 1: 2017 SDR Data Collection Contacting Protocol and Schedule by Starting Mode**

| Week | Web Start Mode<br>Contacts   |   | Mail Start Mode<br>Contacts   | CATI Start Mode<br>Contacts   | Week |
|------|--|---|-------------------------------|-------------------------------|------|
| P    | <i>PREFIELD LOCATING</i>   |   |                               |                               | P    |
| 0    |  |   | Prenotice Letter              | Prenotice Letter              | 0    |
| 1    | Initial Contact Letter   |   | Questionnaire Mailing #1      |                               | 1    |
|      | Initial Contact Email  |   |                               |                               | 2    |
| 2    |  |   | Thank you/Reminder Postcard   |                               | 3    |
| 3    | Follow-up Letter   |   |                               |                               | 4    |
| 4    | Follow-up Email  |   |                               |                               | 5    |
| 5    |  |   | Questionnaire Mailing #2      |                               | 6    |
| 6    | Telephone follow-up  | Questionnaire Mailing #1                          |                               | Prompting Letter w/Web access | 6    |
|      |  | Thank you/Reminder Postcard                       |                               | Prompting Email w/Web access  |      |
| 7    | <i>Web preference sample members</i>   |   | Prompting Letter w/Web access |                               | 7    |
| 8    |  | Telephone follow-up                               | Prompting Email w/Web access  |                               | 8    |
| 9    |  |   |                               | Questionnaire Mailing #1      | 9    |
| 10   | Questionnaire Mailing #1   | <i>Push to Web and Newly Found Sample Members</i> | Telephone prompt              | Thank you/Reminder Postcard   | 10   |
| 11   | Thank you/Reminder Postcard  |   |                               |                               | 11   |
| 12   |  |   | Telephone follow-up           | Survey Request Letter         | 12   |
| 13   | Survey Request Letter  |   |                               | Paired Email                  | 13   |
| 14   | Paired Email   |   |                               |                               | 14   |
| 15   | Determine Late-Stage Priority Score; Isolate and select sample for Late-Stage treatments;        |   |                               |                               | 15   |
| 16   | Prepare Late-Stage mailings; Sample rests  |   |                               |                               | 16   |
| 17   | LATE-STAGE PHASE<br>Late-Stage Incentive Offer and Contacting Protocol                           |   |                               |                               | 17   |
| 18   |  |   |                               |                               | 18   |
| 19   |  |   |                               |                               | 19   |
| 20   |  |   |                               |                               | 20   |
| 21   |  |   |                               |                               | 21   |
| 22   | Isolate and select sample for Last Chance treatments; Prepare Last Chance mailings; Sample rests |   |                               |                               | 22   |
| 23   | LAST CHANCE PHASE<br>and continue to find and offer Late-Stage Incentive                         |   |                               |                               | 23   |
| 24   |  |   |                               |                               | 24   |
| 25   |  |   |                               |                               | 25   |
| 26   |  |   |                               |                               | 26   |
| 27   |  |   |                               |                               | 27   |
| 28   |  |   |                               |                               | 28   |
| 29   | FINAL REQUEST PHASE  |   |                               |                               | 29   |
| 30   |  |   |                               |                               | 30   |

**3.4 Incentive Plan for 2017**

The 2017 SDR protocol includes an early and a late-stage incentive for U.S.-residing nonrespondents to reduce nonresponse bias. Sample members determined to be out of the U.S. or who work for the National Science Foundation will be excluded from the incentive offer, even if the contacting information on file for them is residential.

As noted in section A.9, there are two primary sample components in the 2017 SDR:

1. Panel: Individuals included in the 2015 SDR sample and selected for the 2017.
2. New cohort: Individuals who received their doctorate in the academic years 2014 and 2015.



In the 2017 SDR, these specific subgroups of cases will be offered an early incentive:

- **Reluctant Panel Sample Members.** Sample members who completed the 2006, 2008, 2010, 2013, or 2015 SDR only after having been offered an incentive are designated as reluctant, and will receive a \$30 incentive check affixed to the cover letter of the first mail questionnaire; the follow-up letter to these sample members will refer to this incentive. The rationale for proposing this approach is based on the 2013 SDR data collection experience – 69.7% of the incentivized “incentive required” sample members completed the survey compared to 37.6% of the non-incentivized “incentive required” sample members who completed the survey. In 2015, all sample members who only responded after receiving an incentive were sent an incentive with their first survey request. Of the cases from this group, 81.5% completed the 2015 survey.
- **New Cohort Sample Members.** Based on the new cohort incentive experiments in the 2006 and 2008 SDR, an incentive will be included in the second contact with all new cohort sample members. The 2006 and 2008 experiment results indicate that offering an incentive in the second request for survey participation was more effective than offering it in the first survey request or during the late-stage of data collection. These experimental results suggest an incentive offer to new cohort sample members accelerates their response and will be more cost-effective. These sample members will receive a \$30 incentive check in their second contact regardless of their starting mode (although for sample members starting on the telephone, this will be their first mailing, which follows their first “contact” of a telephone call) – only finalized cases (i.e., completes, final ineligible, final refusals) will not receive the second contact with the incentive.

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 midway through the field period will have a probability of receiving a monetary incentive. In the late-stage incentive plans used for the 2008 through 2015 SDR, a higher probability of selection for the incentive was given to more challenging cases in key analytic domains with relatively lower response rates, in order to improve the accuracy of survey estimates and, ideally, mitigate nonresponse bias. Preliminary results from the 2015 SDR, show that found late-stage eligible cases offered the incentive achieved a survey yield of 56.5% versus late-stage incentive eligible cases not offered the incentive who achieved a survey yield of 51.5%. Based on these results and findings from past cycles, we propose to continue that strategy for the 2017 cycle.

To effectively allocate limited resources for the monetary incentive to late-stage survey nonrespondents, there will be an analysis of the characteristics of the remaining nonrespondents using a logistic regression model and/or a Mahalanobis distance measure to determine which types of sample members should receive additional inducement to mitigate response bias; the cases with lowest response propensity and/or who are least similar from existing survey participants will be selected for the incentive provided they reside in the U.S. The volume of late-stage nonresponse cases to be incentivized will be based on available funds.

Also during late-stage, any locatable nonresponding sample members selected for any early incentive who were not previously sent their incentive due to locating problems or a lack of a mailing address, will be issued or reoffered the incentive at this time. Those nonrespondents who were successfully sent the incentive during the early phase will receive the non-incentivized late-stage treatment.

Based on this monetary incentive plan which includes early and late-stage incentives, the incentive will be offered to approximately 24,600 sample members to include an estimated 19,000 panel and 5,600 new cohort sample. All SDR incentive experiments have consistently shown that most incentivized sample members do not cash their prepaid incentive check, yet do participate in the survey. For example, in the 2015 SDR, 23,738 sample members were offered the incentive. Of these individuals, only 8,550 cashed the incentive check (36.0%) and yet 15,757 completed the survey (66.3%). Therefore, although the incentive to be offered to the approximately 24,600

sample members in the 2017 SDR will have a total incentive amount of \$738,000, the actual value of the cashed incentives is expected to total \$266,000.

#### **4. TESTING OF PROCEDURES**

Data from both SDR and NSCG are combined into a unified data system. Therefore, the two surveys must be closely coordinated to provide comparable data. Many of the questionnaire items in the two surveys are the same.

The integrated 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 SDR or NSCG surveys. These items are essential for sampling, respondent verification, basic labor force information, or analyses of the science and engineering workforce in the NCSES integrated data system. SDR and NSCG surveys ask core questions of all respondents each time they are surveyed to establish the baseline data and to update the respondents' labor force status, changes in employment, and other characteristics. Module items are special topics that are asked less frequently on a rotational basis. Module items provide the data needed to satisfy specific policy or research needs.

Beginning in 2017, there will be only one SDR questionnaire for all sample members, regardless of their current location (in the U.S. versus out of the U.S.). The 2017 questionnaire contains no new survey question content compared to the 2015 version of the questionnaire, though minor response category adjustments were made to two items to accommodate persons living out of the country. Specifically, under Question A13 "Which one of the following best describes your principal employer during the week of February 1, 2017?" a category was added to identify those employed by a non-U.S. government. Also, question E9, which follows a skip pattern for those who were non-U.S. citizens on the reference date, now includes a third category for those doctorate recipients who no longer held a U.S. resident visa on the reference date in addition to the other two categories of "With a permanent U.S. Resident Visa", and "With a temporary U.S. Resident Visa."

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

Survey contact materials will be tailored to fit sample member's information and to gain their cooperation. Contact materials that request sample member participation via the web survey will include access to the survey online. As has been done since 2003 SDR, the 2017 SDR letterhead stationery will include project and NSF/NCSES website information, and the data collection contractor's project toll-free telephone line, USPS and email addresses. Stationery will contain a watermark that shows the survey's established logo as part of an effort to brand the communication to sample members for ease of recognition. The back of the stationery will display the basic elements of informed consent.

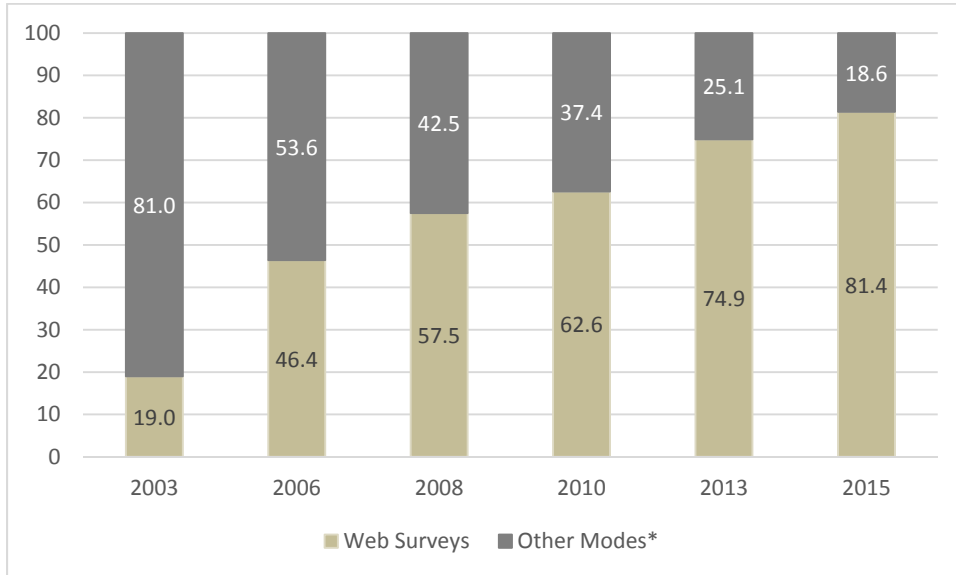
##### **4.2 Questionnaire Layout**

Other than minor question adjustments described above, there are no changes for 2017. Through the Human Resources Experts Panel (section A.8), cognitive research and testing, and other community interest, NCSES continues to review and revise the content of its survey instruments. NCSES will review the data after the 2017 round, and will propose and test changes for the 2019 questionnaire.

### 4.3 Web-Based Survey Instrument

In the 2003 SDR, the online mode was introduced. Figure 2 shows the rate of SDR web survey participation from the 2003 through 2015 survey cycles.

**Figure 2: Web Mode Participation Rate: 2003-2015 SDR**



\*Other response modes are self-administered mail-in form or telephone interview.

As in 2015, the 2017 online survey will employ a front-end user interface that is optimized for mobile devices (e.g., smartphones and tablets) so that the respondent experience with the online survey will be similar regardless of the screen size or web browser used to access the survey. Over 80% of the SDR respondents are expected to participate via web based on their stated preference in the last round and the observed rate of online participation in the last survey cycle (81% in 2015).

## 5. ADAPTIVE DESIGN GOALS, MONITORING METRICS AND NONRESPONSE ERROR ASSESSMENT

Like the 2015 SDR, the 2017 cycle will follow an adaptive design strategy. This is only the 2<sup>nd</sup> cycle of SDR to apply an adaptive design approach and the emphasis is to establish a data quality monitoring system to assess the effects of different locating and data collection practices. The adaptive design goals are not necessarily to maximize the survey's overall response rate but rather to achieve a balanced sample to minimize potential nonresponse bias and to obtain sufficient responses to support small domain estimation. Surveys often try to correct for nonresponse bias after data collection using weighting, post-stratification, or other adjustments. Adaptive design strategies attempt to correct for nonresponse bias during data collection by changing the respondent population to be more balanced on frame characteristics related to response and outcome measures.

Monitoring metrics used in the SDR adaptive design approach to achieve a balanced sample include R-indicators and Mahalanobis distance measures. As a metric, R-indicators are useful for measuring overall response balance and to identify *subgroups* that should be targeted to increase response balance. Another metric, the Mahalanobis distance measure, identifies *specific cases* in those subgroups that also are likely to have an effect on nonresponse bias, and thus are the optimal priority cases for intervention, both from a response balance and nonresponse bias perspective. These monitoring metrics will direct the allocation of data collection resources to the more influential cases and cases that will contribute the most to coverage of small domains.

As noted in section 3.3 and shown in Figure 3, the SDR data collection field period is segmented into four phases: a “starting” phase, “interim” phase, “late-stage” phase, and “last chance” phase. Prior to the start of each data collection phase, sampled cases will be prioritized per the importance of their representation in the 2017 sample. Prioritization of the 2015 sampled cases for the starting stage in the 2017 cycle will be based on several factors including their 2015 priority score, and the 2015 level of effort in locating and contacting the respondent. Prioritizing the new cohort sampled cases will be based largely on FFOD domain size at the start of data collection. FFODs with small population counts will be given higher priority.

The late-stage incentive offer and appropriate follow-up responses will vary for high and low priority cases under the adaptive design approach and its monitoring metrics. Figure 3 shows how locating and data collection treatments will vary for high and low priority cases throughout the field period. It is not expected that this adaptive design strategy with differential locating and data collection treatments will eliminate all survey nonresponse and potential response bias, but it is expected that this strategy will help mitigate bias and further minimize total survey error.

At the conclusion of survey data collection, an assessment of potential bias in the survey results will be conducted. Numerous metrics will be computed to assess bias: unit response rates, estimates of key domains, item nonresponse rates, and R-indicators. Each of these metrics provides different insights into the issue of nonresponse.

Unit response rates quantify the percentage of the sample population that responded to the survey. In the 2015 SDR, the sample had an overall weighted response rate of 66.0%; however, the weighted response rates by fine field of degree which defined the strata ranged from 50.5 to 91.0%. Some variation in response is expected due to random variation; large variations in response behavior can be a cause for concern with the potential to introduce nonresponse bias.

An examination of the estimates of key analytic domains provides insight on the potential for bias due to nonresponse error and the impact on the survey estimates. To account for nonresponse, and ensure the respondent population represents the target population in size, nonresponse weighting adjustments are made to the respondent population. Following the nonresponse adjustment, post-stratification is employed to ensure the respondent population represents not just the size of the target population, but also the proportion of members in various domains of the population. To estimate the effect of these adjustment steps, estimates of various domains within the SDR target population will be calculated from the frame, from respondents, after the nonresponse adjustment, and after final adjustments. This examination will provide insight on whether the SDR weighting adjustments are appropriately meeting the SDR survey estimation goals.

Like the unit response rates, the item response rates can be used as an indicator for potential bias in survey estimates. To examine item nonresponse, response rates for all questionnaire items will be produced. In addition, to examine the impact of data collection mode on item nonresponse, item response rates by response mode also will be produced.

**Figure 3: 2017 SDR Data Collection and Adaptive Design Overview**

| Week | Locating Activity   | Contacting Protocol  |  |                          | Adaptive Design Treatments Based on Assessed Priority |  |  |   |
|------|---|--|--|--------------------------|---|--|--|---|
|      |   | Web Start Mode Contacts  | Mail Start Mode Contacts                   | CATI Start Mode Contacts | Phase   | Locating   | Data Collection  |   |
| 0    | <p>When cases are found, they receive data collection outreach in real-time via the case management system;</p> <p>cases found during the Starting and Interim Phase join the Web Start Mode protocol.</p> <p>Cases determined to have outdated contacting information in data collection are sent to locating in real-time via automated case management system rules.</p> |  | Prenotice Letter                           | Prenotice Letter         | Starting  | Cases searched in priority order.  | No differential treatment  |   |
| 1    |   | Initial Contact Letter   | Mail Questionnaire                         |                          |   | <ul style="list-style-type: none"> <li>• AIM locating treatment</li> <li>• Limit to 60 minutes per case</li> </ul> <u>Low Priority Treatment:</u> <ul style="list-style-type: none"> <li>• Limit to 20 minutes per case</li> </ul> |  |   |
| 2    |   | Paired Email   | Postcard mailing                           | CATI calling             |   |  |  |   |
| 3    |   | Follow-up Letter   |  |                          |   |  |  |   |
| 4    |   | Paired Email   |  |                          |   |  |  |   |
| 5    |   |  | Remail Quex                                |                          |   |  |  |   |
| 6    |   | Telephone prompt   | Mail Questionnaire                         |                          | Prompting Letter                                      | Interim  | Cases searched in priority order.  | No differential treatment   |
| 7    |   | Web preference sample members                                      | Postcard mailing                           |                          | Paired Email  |  | <p>Same treatment as in Starting Phase plus...</p> <p><u>High Priority Treatment adds:</u></p> <ul style="list-style-type: none"> <li>• Expert locating</li> </ul> <p><u>Low Priority Treatment remains:</u></p> <ul style="list-style-type: none"> <li>• Basic search protocol</li> </ul> |   |
| 8    |   |  |  | Prompting Letter         |   |  |  |   |
| 9    |   |  | Telephone prompt                           | Paired Email             |   |  |  |   |
| 10   |   | Mail Questionnaire   | Push to Web and Newly Found Sample Members | Telephone prompt         | Postcard mailing                                      |  |  |   |
| 11   |   | Postcard mailing   |  |                          | Mail Questionnaire                                    |  |  |   |
| 12   |   |  |  | Telephone follow-up      |   |  |  |   |
| 13   |   | Survey Request Letter  |  |                          | Survey Request Letter                                 |  |  |   |
| 14   |   | Paired Email   |  |                          | Paired Email  |  |  |   |
| 15   |   | Update Priority Score and Prepare for Outreach                     |  |                          |   |  |  |   |
| 16   |   | *** All nonresponse cases combined regardless of starting mode *** |  |                          |   |  |  |   |
| 17   |   | High Priority  |  |                          | Low Priority  | Late-Stage   | Same High and Low treatments as Interim Phase, but cases may change priority order.  | <p><u>High Priority Treatment</u></p> <ul style="list-style-type: none"> <li>• Questionnaire with...                             <ul style="list-style-type: none"> <li>- \$30 offer (US) or</li> <li>- Infographic (non-US)</li> </ul> </li> </ul> <p><u>Low Priority Treatment:</u></p> <ul style="list-style-type: none"> <li>• Questionnaire (US) or Infographic letter (non-US)</li> </ul> |
| 18   |   | Mail Quex w/\$30 (US) or Infographic (non-US)                      |  |                          | Mail Questionnaire (US)                               |  |  |   |
| 19   |   | Paired Email   |  |                          | Paired Email  |  |  |   |
| 20   |   | Telephone follow-up  |  |                          | Telephone prompt                                      |  |  |   |
| 21   |   |  |  |                          |   |  |  |   |
| 22   |   | Update Priority Score and Prepare for Outreach                     |  |                          |   |  |  |   |
| 23   |   | High Priority  |  |                          | Low Priority  | Last Chance  | Same High and Low treatments as Interim Phase, but cases may change priority order.  | <p><u>High Priority Treatment:</u></p> <ul style="list-style-type: none"> <li>• Higher level of effort (more contacts)</li> </ul> <p><u>Low Priority Treatment:</u></p> <ul style="list-style-type: none"> <li>• Lower level of effort and CIO offered sooner (fewer contacts)</li> </ul>   |
| 24   |   | Infographic Letter   |  |                          | CIO Letter  |  |  |   |
| 25   |   | Paired Email   |  |                          | Paired Email  |  |  |   |
| 26   |   | Telephone prompt   |  |                          | Telephone prompt                                      |  |  |   |
| 27   |   |  |  |                          |   |  |  |   |
| 28   |   | CIO Letter / Paired Email  |  |                          | CIO Letter or Email                                   |  |  |   |
| 29   |   | Telephone prompt   |  |                          |   |  |  |   |
| 30   | Final Letter / Paired Email   |  |  |                          |   |  |  |   |

AIM = Accurant Individual Match search. CATI = Computer-assisted telephone interview. CIO = Critical Item Only survey version.

## **6. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION**

Chief consultants on statistical aspects of data collection at NORC at the University of Chicago are Donsig Jang, Director of the Center for Excellence in Survey Research (301-634-9415) and Michael Yang, Senior Statistician (301-634-9492).

At NCSES, the contacts for statistical aspects of data collection are Samson Adeshiyan, NCSES Chief Statistician (703-292-7769), Daniel Foley, SDR Project Officer (703-292-7811) and Wan-Ying Chang, NCSES Mathematical Statistician (703-292-2310).