

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Respondent Universe and Sampling Methods

The 2015 SDR features a substantial sample expansion, and significantly expanded analytical objectives. The SDR sample size is increasing from 47,078 cases in 2013 to 120,000 cases in 2015. The main objective of this expansion is to produce reliable estimates of employment outcomes by the fine field of degree (FFOD) taxonomy used in the Survey of Earned Doctorates (SED). The expanded sample will maintain the existing estimation capabilities associated with analytical domains defined by various demographic characteristics and currently used in NCSES publications such as *Science and Engineering Indicators*, and *Women, Minorities, and People with Disabilities*.

Prior to the 2015 sample expansion, the SDR sample for each cycle consisted of two components: a panel sample and a new cohort sample. The panel sample was selected from the previous cycle's SDR sample, and the new cohort sample was selected from an SED frame representing the most recent two academic years of science, engineering, and health research doctorate recipients..

The 2015 SDR is comprised of three components: 1) a panel sample that tracks doctorate recipients throughout their careers until age 76 referred to as the "panel," 2) a new sample of doctorate recipients awarded their degrees from 1959 to 2011 referred to as the "expansion," and 3) a new cohort component that adds new doctorate recipients after they receive their degree (in the 2012 and 2013 academic year) referred to as the "new cohort." The 2015 panel will carry over about one third of the 2013 panel cases.

B.1.1 Frame

The 2015 SDR sampling frame is based on the most current Doctorate Record File (DRF). The DRF is a cumulative census of research doctorates awarded from U.S. institutions since 1920. It is annually updated with new research doctorate recipients through the SED. The 2013 DRF contains 1,995,942 records.

To be eligible for inclusion in the 2015 SDR sampling frame, an individual in the DRF must:

- Have a research doctoral degree in a science, engineering or health field from a U.S. institution, awarded from academic years 1959 through 2013, and
- Be non-institutionalized and under age 76 as of February 1, 2015.

The final 2015 SDR sampling frame consists of 1,102,985 cases classified into six groups based on the new SDR design, and is shown in Table 1. The "Sample Component" column refers to the three components described above: 1) one group for the panel; 2) four groups for the expansion sample; and 3) one group for the new cohort of recent degree recipients. The six frame groups can be described as follows:

1. Frame Group 1 contains the eligible cases from the 2013 SDR selected sample and represents a longitudinal panel.
2. Frame Group 2 contains cases previously selected for the SDR and who were determined to be permanently ineligible during past survey data collection. The majority of these cases are verified as deceased or are individuals who never earned a science, engineering, or health doctorate degree. Because this information is known for cases included in past SDR cycles and is not available for frame cases never selected SDR in the past, these known ineligible cases are included in the frame and given a chance at selection. The purpose of including these cases on the

frame is to simplify the weighting process because these known ineligible cases will have to be accounted for during estimation. By including them on the frame, the survey will not need to track or estimate these cases separately. Any of these cases selected will not be fielded, but will retain their permanent ineligible status for response calculations for the 2015 SDR.

3. Frame Group 3 contains cases previously selected for the SDR, but were excluded from the sample in a maintenance cut procedure. A proportional maintenance cut of panel cases was implemented in the past to control the overall sample size while adding in representative cases selected from the newest cohorts of doctorate degree earners.
4. Frame Group 4 is the largest group and contains cases that were eligible for the SDR in prior cycles, were given a chance at selection for the survey, but were never selected.
5. Frame Group 5 contains cases that were never given a chance at selection for the SDR due to eligibility rules in effect at the time of sampling. Specifically, for the 1977 through 2001 SDR survey cycles, individuals who earned a science, engineering or health research doctorate degree from a U.S. institution who were not U.S. citizens and reported plans to leave the U.S. subsequent to degree award were not considered eligible; these cases were excluded from the 1977 to 2001 SDR sample frames. Beginning with the 2003 SDR cycle and all subsequent survey cycles, the sample included these types of individuals.
6. Frame Group 6 contains individuals who have earned their degree since the 2013 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 2012 and 2013 academic years.

Table 1: The Six Groups of 2015 SDR Frame Cases

Sample Component	Frame Group	Description	SED Academic Years (AY)	Frame Cases
1 – Panel	1	2013 SDR sampled cases that remain eligible for 2015	1960-2011	45,936
2 - Expansion	2	Permanently ineligible cases determined in past cycles of the SDR accrued since 1973 forward (i.e., deceased, no degree earned, non-U.S. citizens located abroad 2 cycles in a row)	1964-2011	2,292
	3	Maintenance cut cases removed from the sample during 1995-2013 sample selection (proportionally deselected regardless of response outcome)	1959-2009	64,532
	4	Eligible for primary selection from SED AY 1959-2011, but not selected previously	1959-2011	859,891
	5	Not eligible for primary selection from SED AY 1975-2000 that are now considered eligible (i.e., new graduates with plans to leave the U.S. after degree award)	1975-2000	52,811
3 – New Cohort	6	New cohort cases from SED AY 2012 and 2013	2012-2013	77,523
Total				1,102,985

B.1.2 2015 Sample Design

The SDR has historically featured a stratified random sample design, where the strata were defined by broad degree field, gender, race, ethnicity, citizenship, disability status, and other relevant demographic variables. The 2013 SDR featured a total of 194 strata, including 150 strata for individuals residing in the

U.S. and 44 strata for individuals residing outside the U.S. The strata for those residing in the U.S. were defined by degree field, gender, race and ethnicity, citizenship at birth, and disability status; strata for those residing outside the U.S. were defined by degree field, gender, race, ethnicity, and citizenship at birth.

The 2015 SDR still has a stratified random sample design but now features a total of 216 newly defined strata. The strata are defined by the SED FOD alone, reflecting the emphasis on the new analytical objectives at the fine field level. 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 (i.e., discontinued fields in the 21st century). 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 118,916 cases were allocated to the 215 strata that constitute the fine fields; 1,084 cases were allocated to the 216th stratum representing the discontinued fields after the 2000 SED.

The 2015 SDR design also features oversampling of under-represented minorities (URM), women, and the 2013 panel cases within each sampling stratum. Oversampling of URM and women allows the sample to sustain the estimation capabilities under the prior SDR design. The purpose of a panel oversample is to support limited but important longitudinal analysis using earlier waves of the SDR. The addition of a panel oversample maintains the simplicity of a new sample, but allows for longitudinal analysis using earlier waves of the SDR. 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 2015 SDR sample allocation follows a two-step process. 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.

For 2013, the overall design effect was 1.13, reflecting mostly proportional sample allocation under that design. The projected overall design effect for 2015 increases to 1.6 due to a more disproportional sample allocation under the new design. 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.

B.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.

B.2.1 Weighting

A final weight will be computed for each completed interview to remove 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 2015 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 2013 SDR. First, for cases whose eligibility status is not determined by the end of 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 poststratification adjustment align 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. For example, the 2013 SDR is poststratified by degree year and the 15-category broad degree field.

For the 2015 SDR, some modifications to the weighting scheme may be necessary to reflect the changes to the sample design. In 2013, logistic regression models were used to derive unknown eligibility and nonresponse weighting adjustment factors for different segments of the sample. The 2015 weighting will take the same approach, although the actual models have not yet been developed and will likely be different from the models used in 2013. With a final weight, the Horvitz-Thompson estimator will be used to derive point estimates for various SDR variables.

B.2.2 Item Nonresponse Adjustment

The SDR has conducted comprehensive imputation to fill in item-level missing data. Two general methods of imputation, logical imputation and statistical imputation, have been used. The logical imputation method is usually employed during the data editing process when a possible 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, the statistical imputation is employed and uses primarily a hot-deck imputation procedure. 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 considered to be similar to the donor respondent. The 2015 SDR will use similar imputation techniques, although the actual imputation models may be different to reflect changes in the sampling frame and the final sample.

B.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 2015 SDR, which employs systematic sample after sorting cases within each stratum by selected demographic variables. Under the SDRM method, implicit strata imposed by the sorting are collapsed and averaged over all possible pairs so that variance of an implicit stratum can be estimated. In 2013, a total of 104 replicates were developed and the final weight was adjusted within each replicate such that standard replication techniques could be used to estimate the SDRM variance. The same general approach will be adapted to the 2015 SDR.

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. Users who do not have access to such software will be provided generalized variance functions (GVFs) to support

variance estimation. The GVs will be estimated for the key SDR domains and instructions will be provided on how to use the GVs to estimate the variance of various sample statistics and for domains not directly covered by the GVs.

B.3 Methods to Maximize Response

B.3.1 Maximizing Response Rates

The weighted response rate for the 2013 SDR was 76.4 percent. Extensive locating efforts, nonresponse follow-up survey procedures, and targeted data collection protocols will be used to attain a targeted 70 percent response rate for 2015. 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.

B.3.2 Locating

New sample members will first need to be located before a request for survey participation can be made. The 2015 SDR will follow a locating protocol similar to that implemented in 2013 with some improvements. The contacting information obtained from the 2013 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 cases that are new-to-SDR³ in 2015.

2015 SDR Locating Protocol Overview. As in the prior SDR cycles, there will be two phases of locating for the 2015 SDR: prefield and main locating. Prefield locating activities include Accurint®⁴ batch processing, 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 work does not include direct contact (by mail, telephone, or email) with the sample members or other individuals who many know how to reach the sample members. Main locating includes manual locating and additional Accurint® processing as needed. The 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 individuals who may know how to reach the sample members. Both the prefield and main locating activities will be supported by an online case management system. The case management system will include background information for each case, all of 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 2015 SDR will implement a responsive 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. All SDR prefield locating activities will be “stealth” meaning the sample members and any individuals who may know the sample member will not be contacted to obtain or confirm up-to-date contacting information.

³ For the 2015 cycle, “new-to-SDR” refers to cases not included in the 2013 sample due to the expansion and new cohort sample components (components 2 and 3, respectively) described in the Section B.1.1 Frame.

⁴ 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.

1. For the panel (sample component 1) and the new cohort (sample component 3), 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 will be more than 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 expansion and new cohort sample components. Prefield locating will be conducted on panel cases which could not be found in the prior round of data collection.

For the new-to-SDR cases (the expansion and new cohort sample components), 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:

- 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 as part of the manual locating effort throughout prefield locating, for those individuals not found via the automated searches. Only publically 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 2015 SDR will use Accurant® 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 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, but are now able to call to confirm the correct sample member has been reached and, ideally, to gain survey cooperation from the found individuals. In addition to outreach to the sample members, main locating activities during data collection will include calls, letters, and emails to dissertation advisors, employers, alumni associations, and other individuals who may know how to reach the sample member.

B.3.3 Data Collection Strategies

A multi-mode data collection protocol (mail, CATI, and web) will be used to facilitate survey participation, data completeness, and sample member satisfaction. The 2015 SDR data collection protocols and contacting methods are built upon the methodology used in the 2013 and prior survey cycles incorporating improvements based upon the methodological experiment results from the 2013 cycle (see Attachment G - 2013 Survey of Doctorate Recipients Contacting Protocol Experiments Results). The data collection field period is segmented into two phases: a "starting" phase and "late-stage" phase. The starting phase includes four separate data collection protocols tailored to different sample groups. In the late-stage, all remaining nonresponse cases (regardless of their starting data collection protocol) receive a single contacting protocol.

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 – The initial request to complete the 2015 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:
 - Panel members that report a preference for the online mode.
 - Panel members that report a preference for the mail or CATI mode, but who have both a USPS and email address. This assignment is based on the results from a 2013 contacting experiment which indicated sample members can be converted to the online mode when provided a reasonable rationale for using the online mode.

- Expansion sample members with a USPS mailing address and/or email address. In 2013 SDR, 66.7 percent of those reporting a mode preference selected the web or online mode for future survey rounds, and 74.9 percent of the 2013 SDR surveys were collected in the online mode.
 - New cohort sample members who have a USPS mailing address and a portable email address not tied to their doctoral degree institution. A 2013 contacting experiment indicated new cohort sample members have a higher response and respond more quickly when started in the online mode.
2. Mail – The initial request to complete the 2015 SDR will be made through a USPS mailing that includes a paper version of the survey. This starting protocol group includes 2013 panel sample members who reported they prefer the mail mode and do not have an email address, and 2013 respondents who completed by CATI without a reported survey mode preference and without an email address. 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 2015 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, as well panel sample members who refused to participate in 2013.
 4. CATI – The initial request to complete the 2015 SDR will be made by a trained telephone interviewer who will attempt to complete the survey via CATI. This starting group includes 2013 panel sample members who reported they prefer the CATI mode and other sample members from any other sample component 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 D – Draft 2015 SDR Survey 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 primarily based on type of cohort (e.g., 2013 panel, expansion, or new cohort). Additional tailoring for the 2013 panel members will be 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 2013 cycle, but with some improvements described below.

1. In the 2015 cycle, the 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 the survey is shorter.

2. In 2013, 114 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, as has been the case in previous cycles. However, in 2015, the contact materials will revert back to text used in the 2010 cycle which is less assertive than the text used in 2013. In 2010, seven sample members participated after receiving the mailing and none complained. In 2013, more assertive text only yielded two complete surveys and 15 complaints.
3. As noted above, the expansion sample component will be assigned to the web starting protocol. Research⁵ shows that including a 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 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 panel will not mention the survey, but will be graphically appealing and will include useful information like measurement conversions.
4. Incentives will be offered during the starting data collection phase as described below in section B.3.4

CATI 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 multimode effort, the CATI system 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 CATI Interviewing team will include Refusal Avoidance and Conversion specialists who have a proven ability to work with doctoral sample members to obtain informed consent and survey participation.

The overall 2015 SDR schedule of contacts by starting protocol is shown in Figure 1.

⁵ 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: 2015 SDR Data Collection Contacting Protocol and Schedule by Starting Mode

Week	Web Starting Protocol	Mail Starting Protocol	Reluctant Mail Starting Protocol	CATI Starting Protocol	Week
	Includes: • Panel members except past refusers, and those who prefer mail or CATI w/o an email address; • Expansion sample members; • New cohort sample members with portable email addresses	Includes: • Panel members who prefer mail w/o an email address; • Panel members who completed by CATI w/o a mode preference and w/o an email address • New cohort members w/o a portable email address	Includes: • Panel members who will only participate with an incentive • Panel members who refused to participate in 2013	Includes: • Panel members who prefer CATI w/o an email address • Any sample members whose only contact info is a phone number	
0	Initial Contact Letter w/Web access	Prenotice Letter		Prenotice Letter	0
1	Initial Contact Email w/Web access	Questionnaire Mailing #1	Questionnaire Mailing #1	CATI calling	1
2	Follow-up Letter w/Web access	Thank you/Reminder Postcard	Thank you/Reminder Postcard		2
3	Follow-up Email w/Web access				3
4					4
5	Telephone follow-up	Questionnaire Mailing #2	Prompting Letter w/Web access		5
6			Prompting Email w/Web access	Prompting Letter w/Web access Prompting Email w/Web access	6
7		Prompting Letter w/Web access			7
8	Questionnaire Mailing #1	Prompting Email w/Web access		Questionnaire Mailing #1	8
9	Thank you/Reminder Postcard	Telephone prompt	Telephone prompt ³	Thank you/Reminder Postcard	9
10		Telephone follow-up			10
11					11
12	Questionnaire Mailing #2				Questionnaire Mailing #2
13	Isolate and select Late-Stage Incentive sample; Prepare Late-Stage mailings; Sample rests				13
14	Late-Stage Incentive Offer (to those selected) and Contacting Protocol				14
15					15
16					16
17					17
18	Offer Critical Item Only (short) survey to most nonrespondents and continue to find sample members and offer Late-Stage Incentive				18
19					19
20					20
21					21
22	Last Chance Letter and Email				22

B.3.4 Incentive Plan for 2015

The 2015 SDR protocol includes an early and a late-stage incentive for nonrespondents to reduce nonresponse bias.

As noted in section A.9.1, there are three primary sample components in the 2015 SDR:

- (1) Panel: Individuals included in the 2013 SDR sample and selected for the 2015.
- (2) New cohort: Individuals who received their doctorate in the academic years 2012 and 2013.
- (3) Expansion sample: Individuals who were not in the 2013 SDR sample and who received their doctorate in the academic years 1961-2011.

In the 2015 SDR, these specific subgroups of cases will be offered an early incentive:

1. Reluctant Panel Sample Members. Sample members who completed the 2006, 2008, 2010, or 2013 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 percent of the incentivized “incentive required” sample members completed the survey compared to 37.6 percent of the non-incentivized “incentive required” sample members who completed the survey.
2. 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 honorarium 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.
3. Underrepresented Minorities in the Panel and Expansion Sample. Similar to the incentive for new cohort sample members, a \$30 incentive check will be offered in the second contact to sample members in either the panel or the expansion component who are URM.

Based on this incentive plan which includes early and late-stage incentives, the incentive will be offered to approximately 22,250 sample members to include an estimated 2,750 panel, 13,500 expansion, and 6,000 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. The incentive offered to the approximately 22,250 sample members will total to an incentive amount sent of \$667,500. In the 2013 SDR, 10,487 sample members were offered the incentive. Of these individuals, 2,896 cashed the incentive check (27.6%). Of the 2,896 individuals that cashed the incentive check, 2,702 completed the survey (93.3%). Of the 10,487 sample members offered the incentive, 5,062 individuals completed the survey (48.3%) despite not cashing the check.

It is estimated for 2015 that 31.5% of the checks sent will be cashed. As a result, the total cost for incentives is estimated to be \$210,000: \$27,000 to the 2013 panel cohort, \$135,000 to the expansion cohort, and \$48,000 to the new cohort.

B.4 Testing of Procedures

Data from both SESTAT surveys (SDR and NSCG) are combined into a unified data system so the two surveys must be closely coordinated in an effort to provide comparable data. Many of the 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, or analyses of the science and engineering workforce in the SESTAT integrated data system. SESTAT 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.

The 2015 SDR questionnaire contains no new survey question content compared to the 2013 version of the questionnaire. However, as part of the 2015 SESTAT planning effort, NCSES asked the Census Bureau's Center for Survey Measurement to conduct an expert review and cognitive interviews for the full set of NSCG questionnaire items, many of which overlap with the SDR questionnaire items. The expert review and cognitive interviews resulted in minor question wording revisions to select items in the SDR questionnaire.

An improvement to the 2015 paper version of the survey has been implemented based on a review of item-nonresponse in the 2013 cycle. In 2013, of all the critical items, date of birth was missing most often. Furthermore, date of birth is the primary means of sample person verification in the paper survey mode and when it is missing it results in recontacting sample members, which is inefficient and may be bothersome to them. To ensure the correct sample member has been found, particularly in the expansion cohort, and to reduce the need for recontacting sample members, the 2015 paper survey version will include three additional sample person verification questions focused on the sample doctoral degree. These questions have been included in the web and CATI versions of the survey since 2003, and are only new to the paper form. Specifically, in 2015, the paper form will ask sample members to provide the degree field of study, month and year of degree award, and academic institution awarding the degree for their first U.S.-earned SEH doctoral degree.

A complete list of modifications to the 2015 SDR questionnaire is included in Attachment C – Draft 2015 SDR Questionnaires.

B.4.1 Survey Contact Materials

Survey contact materials will be tailored to fit sample members' need for information about the SDR 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 2015 SDR letterhead stationery will include project and NSF 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.

B.4.2 Questionnaire Layout

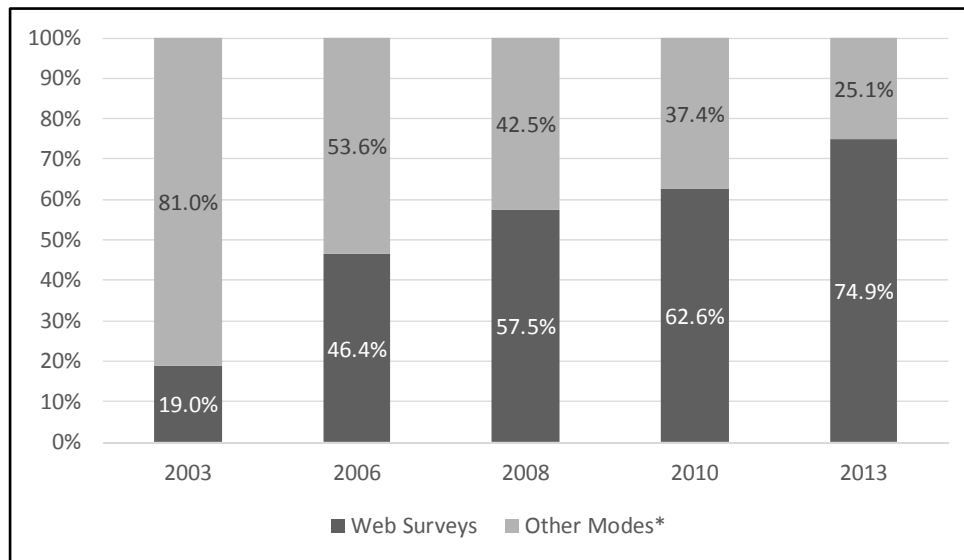
Other than the addition of three sampled degree questions described above, there are no changes for 2015. Through means such as the Human Resources Experts Panel (section A.8.2), as well as cognitive

research and testing, NCSES continues to review, revise, and enhance the content of its survey instruments. NCSES will review the data after the 2015 round, and will propose and test changes for the 2017 questionnaire.

B.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 2013 survey cycles.

Figure 2: Web Mode Participation Rate: 2003-2013 SDR



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

The 2015 SDR will improve upon the web functionality used in the past rounds. While using the same back-end data collection software, the 2015 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. At least 80 percent 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.

B.5 Responsive Design and Nonresponse Error Assessment

The targeted response rate for the 2015 SDR is 70 percent. The reduction in response from previous cycles is based on the change in the sample composition. Most sample members will be newly selected for the SDR (86.6 percent) and will not have current contacting information. The 2015 SDR will implement a responsive 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. It is not expected that these steps will eliminate all survey nonresponse and potential response bias.

The SDR will conduct a nonresponse bias analysis. At the conclusion of the survey, an assessment of potential bias in the survey results will be conducted. Numerous metrics will be computed to assess nonresponse: 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 responded to the survey. In the 2013 SDR, the sample had an overall weighted response rate of 76.4 percent; however, the weighted response rates by major field ranged from 71 to 82 percent. 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 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.

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. Like the unit response rates, the item response rates can be used as an indicator for potential bias in survey estimates.

R-indicators and corresponding standard errors will be provided for the 2015 SDR. R-indicators are useful for assessing the potential for nonresponse bias. R-indicators are based on response propensities calculated using a predetermined balancing model (“balancing propensities”) to provide information on both how different the respondent population is compared to the full sample population, as well as which variables in the predetermined model are driving the variation in nonresponse.

B.6 Contacts for Statistical Aspects of Data Collection

Contacts for statistical aspects of data collection are Jeri Mulrow, Acting NCSES Chief Statistician (703-292-4784), Steven Proudfoot (703-292-4434, SDR Project Officer) and Wan-Ying Chang (703-292-2310), NCSES Mathematical Statistician).