B.   COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1.   RESPONDENT UNIVERSE AND SAMPLING METHODS

The 2019 SDR sample size is set at 120,000 cases including 95,436 returning sample cases from the last survey cycle, 10,000 new cohort cases who are recent doctorate recipients from academic years 2016 and 2017, and a supplemental sample of 14,564 drawn from the 2015 SDR sampling frame to supplement the portion of the returning sample representing the 2015 SDR target population.

Approximately 40,000 of the 2019 SDR sample who participated in the 2015 SDR will be the basis for establishing a longitudinal panel representing the 2015 SDR target population moving forward into the 2021 SDR cycle of data collection. This panel will be weighted and maintained through the 2025 cycle of the SDR to provide longitudinal data for the 10-year time period 2015-2025. (See "Consultation Outside the Agency" within Section A.8 for background information on the development of an SDR sample design that supports longitudinal analysis, and Appendix G for summary reports).

1.1 Frame

The primary sampling frame source for the SDR is the Doctorate Records File (DRF). The DRF is a cumulative file listing research doctorates awarded from U.S. institutions since 1920. It is annually updated with new research doctorate recipients through NCSES's Survey of Earned Doctorates (SED). The 2017 SDR sample selected from the 2015 DRF represented a surviving population of over one million individuals with SEH doctorates who were less than 76 years of age. The 2019 SDR is expected to represent a slightly larger and growing population of nearly 1.2 million SEH doctorate holders from the 2017 DRF including over 80,000 from the two most recent academic years, 2016 and 2017. In total, the 2017 DRF contains 2,214,261 records which includes individuals with research doctorate degrees in a field other than science, engineering or health (SEH), individuals who are age ineligible for the SDR, and deceased individuals.

The target population for the 2019 SDR includes individuals who must:

- Have earned a research doctoral degree in a SEH field from a U.S. institution, awarded no later than academic year 2017, and

- Be less than 76 years of age on 1 February 2019 based on their date of birth, and

- Be living in a noninstitutionalized setting on 1 February 2019

The final 2019 SDR sampling frame can be classified into three groups as shown in Table 1.

1. Frame Group 1 contains individuals eligible for the 2015 SDR target population. These cases were awarded doctorate degrees in academic years 2013 and earlier.

2. Frame Group 2 contains individuals that became newly eligible for inclusion in the 2017 SDR survey cycle. These cases were awarded doctorate degrees in the 2014 and 2015 academic years.

3. Frame Group 3 contains individuals that became newly eligible for inclusion in the 2019 SDR survey cycle. These cases were awarded doctorate degrees in the 2016 and 2017 academic years.

Table 1: The 2019 SDR Frame Groups by Sample Component

Frame Group	Description	SED Academic Years (AY)	Population Size	Sample Component	Sample Size
1	2015 SDR target population that remain eligible for 2019	1960-2013	991,254	2015 SDR returning sample	84,695
				Supplemental sample	14,564
2	2017 SDR newly sampled cases that remain eligible for 2019	2014-2015	82,266	2017 new cohort	10,741
3	New cohort cases from SED AY 2016 and 2017	2016-2017	83,169	2019 new cohort	10,000
Total			1,156,689		120,000

1.2 2019 Sample Design

In the 2015 survey cycle, the SDR sample size increased from 45,000 to 120,000 individuals to improve the estimation capabilities at the fine field of degree (FFOD) level as reported in the SED. Over 200 FFODs served as the explicit sampling strata in both the 2015 and 2017 sample design. For the 2019 survey cycle, three adjustments to the SDR sample design were made based on feedback from SDR stakeholders in combination with evaluations of the reliability and utility of the 2015 and 2017 estimates at the 200+ FFOD stratification levels. (See section A.8 *Technical and analytical support for the SDR longitudinal design*)

- The first adjustment is to the sample selection stratification. The 2015 and 2017 sampling strata of over 200 FFODs are replaced by a set of 308 sampling strata defined by crossing 77 detailed fields of degree (DFOD) with gender (2 categories: male and female) and underrepresented minority (URM) status (2 categories: URM and non-URM). These stratification changes are designed to produce high quality estimates for population subgroups that can be supported by the sample and are aligned with the NCSES taxonomy of disciplines (TOD).

- The second adjustment is to the estimation precision requirements. Instead of setting the precision requirements at the FFOD level as was done in the 2015 and 2017 cycles, the estimation precision requirements are set for three levels of aggregation over the 308 sampling strata. The estimation precision requirements are included in Table 2.

Table 2: The 2019 SDR Overall Precision Requirements

Domain	Margin of Error	Minimum Number of Completes
DFOD	5%	270
DFOD x SEX	6%	190
DFOD x URM	7%	135

In Table 2, the margin of error is two times that of the standard error associated with estimating a population proportion of 50% at the 90% confidence level. The last column shows the estimated required number of completed surveys to achieve the precision requirement per domain.

- The third adjustment is to the returning cohort's retention rule. For the 2019 SDR survey cycle and beyond, sample cases are dropped from subsequent cycles if they did not respond in the preceding two cycles of data collection after entering the SDR. This is a change from previous cycles where all previously sampled cases were carried forward regardless of prior response status. This decision was made in response to the cost inefficiencies of including these nonresponding cases in subsequent survey cycles.

As part of the implementation of this new rule during the 2019 SDR, it is applicable only to sample cases that entered in the 2015 SDR survey cycle. After dropping the cases new to the SDR in 2015 that did not respond in 2015 or 2017 (n=25,066), the sample representation of the 2015 SDR target population will be boosted in the 2019 SDR by including a supplemental sample of 14,564 cases selected from the 2015 DRF. This supplemental sample also serves the purpose of aligning the overall 2019 sample size with the overall sample size goal of 120,000.

The 2015 and 2017 SDR featured oversampling of under-represented minorities and women within each sampling stratum. Oversampling of URM and women allows the SDR sample to sustain the estimation capabilities under the 2013 and prior cycles' SDR design. This feature is ensured under the 2019 sample design. The 2019 design uses gender and URM indicators explicitly to define sampling strata, coupled with estimation precision requirements set at a level to guarantee that women and URM are oversampled at the level similar to past designs. For the new cohort sample and the supplemental sample selection, demographic variables used for stratification in previous sample designs, such as detailed race/ethnicity categories, citizenship at birth, predicted resident location, FFOD, disability indicator, age, 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 new cohort sample was selected systematically with equal probability within each stratum. The supplemental sample is selected via a systematic probability proportional to size (PPS) method. The purpose of using a PPS method is to oversample frame cases that have responded better in the most recent SDR cycles.

The 2019 SDR sample allocation follows an iterative process of gradually adjusting the initial proportional to population size allocation to meet all levels of estimation precision requirements set in Table 2. This set of allocation is considered the expected allocation under the 2019 sample design. The expected allocation at the stratum level is used to determine the stratum level allocation only for the new cohort and the supplemental sample. The eligible returning sample cases are not affected by the expected allocation. The new cohort allocation is assigned by multiplying the expected allocation and the proportion of new cohort on the population for each stratum. For the supplemental sample, the proportional share of the portion that corresponds to the 2015 target population after subtracting the number of returning 2015 old cohort cases is the needed supplement allocation. However, the total needed supplement allocation exceeds the pre-determined sample size of the supplemental sample. As a result, for all strata needing supplement, the final allocation filled at least 80% of the expected proportional allocation.

2.   STATISTICAL PROCEDURES

The SDR statistical data processing procedures have several components including adjusting weights to compensate for the complex sampling design features, imputing to address item nonresponse, and estimating sampling error

2.1 Weighting

A final weight will be computed for each completed interview to reduce potential bias in 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 using methods similar to 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 2017 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 2017 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 2019 SDR will use similar imputation techniques, although the actual imputation models may differ in that longitudinal data from the 2017 cycle may be used to identify donors.

2.3 Variance Estimation

The SDR has historically used the 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 2019 SDR, which employs systematic sampling after sorting cases within each stratum by selected demographic variables. In 2017, a total of 104 replicates were used. Within each replicate, the final weight is developed using the same weighting adjustment procedures applied to the full sample. 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 2019 SDR.

3. METHODS TO MAXIMIZE RESPONSE

3.1 Maximizing Response Rates

The weighted response rate for the 2017 SDR was 70% (unweighted, 69%). Extensive locating efforts, nonresponse follow-up survey procedures, and targeted data collection protocols will be used to attain a

targeted response rate of 75% for 2019. Both an early-stage and late-stage monetary incentive will be offer as outlined in section A.9 above.

3.2 Locating

Panel sample members who are categorized as locating problems in 2019 and new sample members with incomplete contacting data will first need to be located before making a request for survey participation. The 2019 SDR will follow a locating protocol similar to the approach implemented in 2017. The contacting information obtained from the 2017 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 2019.

2019 SDR Locating Protocol Overview. As in prior SDR cycles, there will be two phases of locating for the 2019 SDR: prefield and main locating. Prefield locating activities include Accurint®[1] batch processing and individual searches, address review, and individual case locating (also called manual locating). Prefield locating occurs at least one month 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 interactive (i.e., real time) online case management system (CMS). 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 CMS information will be integrated with survey paradata and monitoring metrics that support an adaptive design approach. See section B.4.4. for more information on the adaptive design plans for 2019.

Prefield Locating Activities. The prefield locating activities consist of four major components as follows:

1. For the returning panel (sample component 1), the new cohort (sample component 2), and the 2015 DRF supplemental sample (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 could be up to three years out of date.

2. After implementing the NCOA updates for the returning panel, new cohort, and supplemental component, the sample will be assessed to determine which cases require prefield locating. This assessment is different for the panel and supplemental cases than for the new cohort sample components.

---

[1] 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.

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). In addition, prefield locating will be conducted on the newly selected 2015 DRF supplemental cases that were selected for the 2019 SDR. 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 2016 and 2017 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 and make limited calls to sample members and outreach to 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 2019 SDR will use Accurint® to conduct individual matched searches, also known as 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 2017 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 2019 data collection outreach activities. Prior to beginning the main locating work, locating staff who worked during the prefield period will receive refresher training that focuses on maintaining sample members' confidentiality particularly when making phone calls, on 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 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 2019 SDR data collection protocols and contacting methods are built upon the methodology used in 2017. 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 in the prior cycle or their past response modes, 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 2019 SDR will be made by a USPS letter 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 SDR 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, especially the supplemental sampled cases, and

   - Cases who previously experienced language problems.

2. Mail – The initial request to complete the 2019 SDR will be made through a USPS mailing that includes a paper version of the survey. This starting protocol group includes returning panel sample members who reported they prefer the mail mode and do not have an email address, and all known 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 2019 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 2017.

4. CATI – The initial request to complete the 2019 SDR will be made by a trained telephone interviewer who will attempt to complete the survey via CATI. This starting group includes returning 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 Appendix E – Draft 2019 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., returning panel or new cohort). Additional tailoring for the returning 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 are consistent with the approach used in the 2017 cycle.

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 2019 SDR schedule of contacts by starting protocol is shown in Figure 1.

Figure 1: 2019 SDR Data Collection Contacting Protocol and Schedule by Starting Mode

Week	Web Start Mode Contacts		Mail Start Mode Contacts	CATI Start Mode Contacts	Week
P	PREFIELD LOCATING				P
0			Prenotice Letter	Prenotice Letter	0
1	Initial Contact Letter / Initial Contact Email		Questionnaire Mailing #1	CATI calling	1
2			Thank you/Reminder Postcard		2
3	Follow-up Letter				3
4	Follow-up Email				4
5			Questionnaire Mailing #2		5
6	Telephone follow-up	Questionnaire Mailing #1 / Thank you/Reminder Postcard		Prompting Letter w/Web access / Prompting Email w/Web access	6
7	*Web preference sample members*		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	*Push to Web and Newly Found Sample Members*	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; Prepare Late-Stage mailings; Sample rests				15
16					16
17	LATE-STAGE PHASE 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					23
24	LAST CHANCE PHASE and continue to find and offer Late-Stage Incentive				24
25					25
26					26
27					27
28					28
29	FINAL REQUEST PHASE				29
30					30

3.4 Incentive Plan for 2019

The 2019 SDR protocol includes an early-stage and a late-stage incentive for U.S.-residing nonrespondents to reduce the potential for nonresponse bias. Sample members determined to be out of the U.S. will be excluded from the incentive offer, even if the contacting information on file for them is residential. In addition, sample members who work for the National Science Foundation will also be excluded from the incentive offer.

As noted in section A.9, there are three primary sample components in the 2019 SDR:

27

1. Returning panel: Individuals included in the 2017 SDR sample who were selected for the 2019 SDR.

2. New cohort: Individuals who received their doctorate in the academic years 2016 and 2017.

3. Supplemental cohort: Individuals who received their doctorate during or prior to 2015 and were newly selected to participate in the 2019 SDR.

In the 2019 SDR, these specific subgroups of cases will be offered an early-stage incentive:

- Reluctant Panel Sample Members. Sample members who completed the 2006, 2008, 2010, 2013, 2015 or 2017 SDR only after having been offered an incentive are designated as reluctant sampled cases 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. In 2013, SDR only had a late-stage incentive offer and those offered the incentive were selected irrespective of their history of receiving an incentive in past cycles. Therefore, some in this group with a history of only participating after having received a past incentive were selected for the 2013 late stage incentive and some were not selected for the incentive. Findings showed that 69.7% of the incentivized sample members in 2013 with a past history of being incentivized completed the survey compared to 37.6% of the non-incentivized sample members in 2013 with a past history of being incentivized. In 2017, all sample members who only responded after receiving an incentive in past cycles were sent an incentive early with their first survey invitation and 81.5% completed the 2017 survey. This practice will be carried forward going into the 2019 cycle.

- 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 ineligibles, 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 2017 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. Results from the 2017 SDR, show that found late-stage eligible cases offered the incentive achieved a survey yield of 56.5% versus late-stage 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 2019 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 pre-set targets of key analytical domains and a logistic regression derived cooperation propensity 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 in underperforming analytical domains 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 the late-stage data collection phase, 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 25,000 sample members to include an estimated 17,500 from the returning panel and supplemental samples and 7,500 from the 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 2017 SDR, 24,085 sample members were offered the incentive. Of these individuals, only 10,078 cashed the incentive check (41.8%) and yet 15,114 completed the survey (62.8%). Therefore, although the incentive to be offered to the approximately 25,000 sample members in the 2019 SDR will have a total incentive amount of $750,000, the actual value of the cashed incentives is expected to total $315,000.

4.   TESTING OF PROCEDURES

The SDR and NSCG are complimentary workforce surveys. Therefore, the two surveys must be closely coordinated to provide comparable data. Many of the questionnaire items in the two surveys are the same including the reference date of 1 February 2019.

The complimentary 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 and NSCG surveys. These items are essential for sampling, respondent verification, basic labor force information, or analyses of the science and engineering workforce using NCSES data. 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.

As in the 2017 SDR, there will be only one 2019 SDR questionnaire for all sample members, regardless of their current location (in the U.S. versus out of the U.S.). The 2019 questionnaire contains no new survey question content compared to the 2017 version of the questionnaire.

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 2019 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

There are no changes from the 2017 SDR questionnaire for the 2019 survey. 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 2019 round and will propose and test changes for the 2021 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 2017 survey cycles.

Figure 2: Web Mode Participation Rate: 2003-2017 SDR
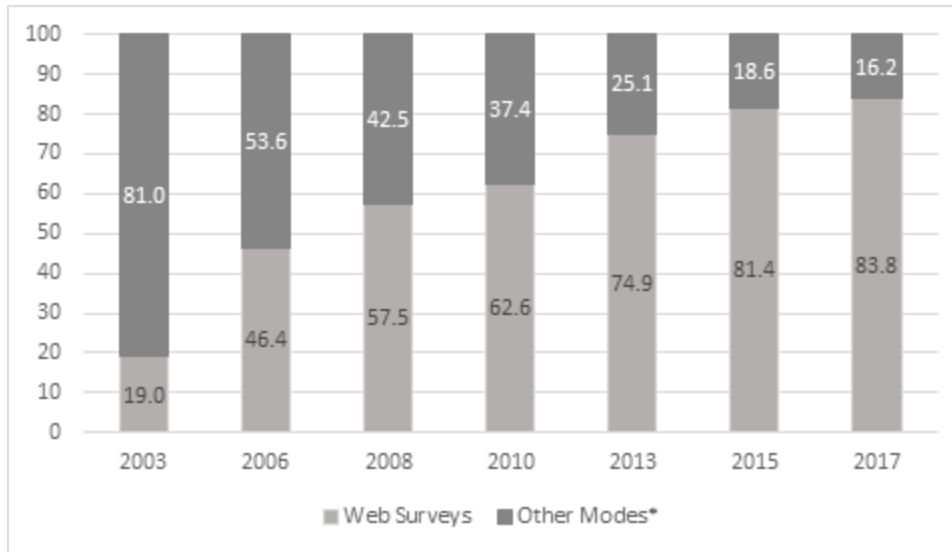
*Other response modes are self-administered mail-in form or telephone interview.

As in 2017, the 2019 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 (83.8% in 2017).

4.4 Adaptive Design Goals and Monitoring Metrics

The 2019 data collection will include an adaptive design strategy to help achieve a balanced sample to minimize nonresponse bias and to achieve target numbers of completes for key analytic domains. This is the 3rd cycle of SDR to apply an adaptive design approach and the 2019 emphasis will be on continuous quality improvement in the use of data monitoring metrics that assess the adaptive design's effects of treatments and interventions used in prioritizing nonrespondents for data collection.

As shown in Figure 3, the 2019 SDR data collection will be implemented in four primary phases, a starting phase, an interim phase, a late-stage phase and a last-chance phase. Prior to the start of each phase, pending nonresponding sampled cases are prioritized with differential locating and/or data collection treatments assigned to each of the priority groups. In general, cases are prioritized according to the importance of their representation in the sample and the estimated effort needed to obtain their

response. The proposed adaptive design metrics used for case prioritization are described below by phase and summarized in Table 3.

For the starting phase in 2019, no data collection treatments occur other than honoring past mode preference among the existing cohort, if applicable. High priority locating treatment may use "people-finder" services and up to 60 minutes of locating effort per case whereas the remaining cases are searched for up to 20 minutes. For nonresponse follow-ups, prioritization of the 2017 SDR sampled cases will be based on their 2017 response status and measures of locating effort defined as the number of changes in locating disposition during the 2017 SDR survey cycle. Prioritizing the new cohort sampled cases and the 2015 supplemental sampled cases will be based on the 2019 SDR prefield locating effort. In 2017, all prior respondents were assigned high locating priority. In addition, unlocated cases and cases receiving lower than average locating effort were also assigned high priority. Within each priority level, cases are sorted by the measure of prior locating effort with those that received lower effort in the previous cycle being processed first. The starting phase of the 2019 SDR data collection generally will last about 4 to 5 weeks. The assignment of high priority treatments is based on the proposed metrics and rules detailed in Table 3.

The interim phase of data collection generally lasts about 10 to 12 weeks. Information from the starting phase response is used to reset the prioritization of all pending nonresponding cases. The number of key analytical domains a sample case is in that are both (a) below the target number of completes and (b) below average response rate will be used to assign high priority treatments to sample cases This combination of criteria is aimed at both helping achieve targets in the analytical domains and improve sample balance by targeting domains that are underperforming. Within each priority level, priority order, which determines the case processing order, is developed using an estimated cooperation propensity score. The cooperation propensity is estimated based on the subset of located cases as of the end of the starting phase using a logistic regression model. Those with higher cooperation propensity are ranked higher for locating, while those with lower cooperation propensity are ranked higher for data collection to ensure that cases needing more data collection effort are worked first. In this phase, high priority for locating is the same as in the starting phase except more experienced locators will be used to perform the locating effort.

The late-stage phase of data collection generally lasts about 6 weeks and the prioritization scheme based on meeting targets in the key analytical domains is continued. To set the priority order, cases are sorted by a cooperation propensity estimated using data as of the end of the interim phase for locating. For data collection, cases are sorted by the number of contacts received, defined as the sum of the number of CATI dials and mailings, and those who received less data collection effort will be processed first. Priority for locating nonrespondents in the late-stage will be the same as in the interim stage (i.e., more locating effort going to high priority cases). In addition, high priority nonrespondents who are found to be residing in the U.S. will be offered a $30 prepaid incentive and an infographic along with a copy of the questionnaire as a data collection treatment to increase the likelihood of gaining their cooperation. The remaining cases who are located in the U.S. only receive a mailed questionnaire or, if residing outside of the U.S., only receive a mailed infographic.

The last-chance phase of data collection generally lasts 6 to 7 weeks. The locating effort is prioritized the same as in the late-stage. However, high priority cases also receive more contacts in attempting to gain their cooperation once found. Furthermore, high priority cases may be offered an abbreviated survey that only asks about "critical items" such as the respondent's employment status, location, citizenship, occupation and employer name. The remaining found sampled members are also offered a "critical item only" (CIO) form. Persons responding to a CIO survey often participate using CATI.

Throughout all phases, other metrics such as R-indicators and preliminary weighted key domain estimates will be monitored and compared to those from prior survey cycles as supplemental metrics to evaluate the impact of the locating and data collection efforts.

Table 3: Proposed 2019 SDR Adaptive Design Metrics and Criteria

Phase	Proposed metric	Proposed metric criteria
Starting (120,000 sample cases; Phase lasts 4-5 weeks)	*Starting phase priority* determined using: - 2017 SDR response status - Measure of prior locating effort defined as the number of changes in locating disposition during the 2017 SDR survey cycle	High priority: All prior respondents and 50% of unlocated cases with fewer than 3 changes in prior locating disposition Note: In 2017, about 75% of overall sample assigned to the high priority treatment. For the 2019 SDR, the size of high priority assignment will be based on level of resources available
Interim (Approximately 80% of original sample cases remain nonrespondents; Phase lasts 10-12 weeks)	*Interim phase priority* determined using - A: The number of key analytical domains a sample case is in that are both (i) below the target number of completes and (ii) below the average response rate - B: Estimated cooperation propensity score based on the subset of located cases as of end of the starting phase	High priority: $A \geq 2$ and $B \leq 0.4$ Locating priority order: Sorting by A (from high to low) and then by B (from high to low) Data collection priority order (CATI prompting order): Sort by A (from high to low) and then by B (from low to high) Note: In 2017, about 75% of pending cases assigned to the high locating treatment
Late-stage (Approximately 50% of original sample cases remain nonrespondents; Phase lasts 6 weeks)	*Late-stage phase priority* determined using: - A1 and A2: The number of key analytical domains a sample case is in that are both (i) below the target number of completes and (A1)/or(A2) (ii) below average response rate - B: Estimated cooperation propensity score based on the subset of located cases as of end of the interim phase - C: Number of contacts received	High priority: A1 ≥ 2 or (A1=1 and A2 ≥ 15) Locating priority order: Sorting by A1 and A2 (from high to low) and then by B (from high to low) Data collection priority order (CATI prompting order): Sort by A1 and A2 (from high to low) and then by C (from low to high) Note: In 2017, about 70% of pending cases assigned to the high priority treatment. High priority nonrespondents who are found to be residing in the U.S. will be offered a $30 incentive (as indicated in Figure 3)

Phase	Proposed metric	Proposed metric criteria
Last-chance (Approximately 35% of original sample cases remain nonrespondents; Phase lasts 6-7 weeks)	*Last-chance phase priority* determined using: - A1 and A2: The number of key analytical domains a sample case is in that are (i) below the target number of completes (A1) or (ii) below average response rate(A2) - B: Estimated cooperation propensity score based on the subset of located cases as of end of the late-stage phase - C: Number of contacts received	High priority: A1 ≥ 3 or (A1=2 and A2 ≥11) Locating priority order: Sorting by A1 and A2 (from high to low) and then by B (from high to low) Data collection priority order (CATI prompting order): Sort by A1 and A2 (from high to low) and then by C (from low to high) Note: In 2017, about 70% of pending cases assigned to the high priority treatment. Located cases may be contacted using Priority mail and may be offered a "Critical Item Only" CATI or web form

Note: The parameters in Table 3 were projected based on the 2017 cycle.

Figure 3: 2019 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			Prenotice Letter	Prenotice Letter	Starting	Cases searched in priority order. High Priority Treatment: • AIM locating treatment • Limit to 60 minutes per case. Low Priority Treatment: • Limit to 20 minutes per case	No differential treatment
1		Initial Contact Letter / Paired Email	Mail Questionnaire	CATI calling			
2			Postcard mailing				
3		Follow-up Letter					
4		Paired Email					
5			Remail Quex		Interim	Cases searched in priority order. Same treatment as in Starting Phase plus... High Priority Treatment adds: • Expert locating. Low Priority Treatment remains: • Basic search protocol	No differential treatment
6		Telephone prompt / Web preference sample members; Mail Questionnaire / Postcard mailing		Prompting Letter / Paired Email			
7			Prompting Letter				
8			Paired Email				
9		Telephone prompt / Push to Web and Newly Found Sample Members		Mail Questionnaire			
10		Mail Questionnaire / Postcard mailing	Telephone prompt	Postcard mailing			
11			Telephone follow-up				
12							
13		Survey Request Letter		Survey Request Letter			
14		Paired Email		Paired Email			
15		Update Priority Score and Prepare for Outreach *** All nonresponse cases combined regardless of starting mode ***					
16							
17		**High Priority** / **Low Priority**			Late-Stage	Same High and Low treatments as Interim Phase, but cases may change priority order.	High Priority Treatment: • Questionnaire with... - $30 offer (US) or - Infographic (non-US). Low Priority Treatment: • Questionnaire (US) or Infographic letter (non-US)
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.	High Priority Treatment: • Higher level of effort (more contacts). Low Priority Treatment: • Lower level of effort and CIO offered sooner (fewer contacts)
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					

Locating Activity (spanning all weeks): When cases are found, they receive data collection outreach in real-time via the case management system; cases found during the Starting and Interim Phase join the Web Start Mode protocol. • Cases determined to have outdated contacting information in data collection are sent to locating in real-time via the case management system. • Cases determined to have outdated contacting information in data collection are sent to locating in real-time via automated case management system rules.

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

34

5. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION

The NCSES contacts for statistical aspects of the SDR 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 and the lead SDR sampling statistician (703-292-2310).