B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Universe and Sampling Procedure

The 2017 ECDS is a survey of individuals who 1) completed their first doctorate degree within the past 10 years, regardless of whether the degree was earned in the U.S. or abroad, and 2) currently work at a U.S. academic institution, federally funded research and development center (FFRDC), or the National Institutes of Health Intramural Research Program (NIH IRP). The primary goal of this survey is to provide: (1) national estimates of the total number of early career doctorates (ECD) in these sectors and (2) robust estimates of key subpopulations of ECD, including the number of postdocs, the number of ECD with foreign-earned degrees, and the number of ECD by sex, citizenship, and race/ethnicity. Currently, data on these subpopulations are extremely limited and the 2017 ECDS sampling procedures have been designed to maximize the precision of these estimates.

The 2017 ECDS includes a two-stage sample design. In the first stage, the sampled institutions will be contacted and asked to provide a listing of ECD working at the institution and, in the second stage, ECD will be sampled and surveyed. The first stage is a stratified sample of institutions—U.S. academic institutions, FFRDCs, and NIH IRP. The set of institutions eligible for the NSF-NIH Survey of Graduate Student and Postdoctorates in Science and Engineering (GSS) will serve as the sampling frame for U.S. academic institutions. While membership in the GSS, FFRDC, and NIH IRP will represent the primary stratification in the institutional sample, the GSS institutions will be further stratified by type of institution defined using Carnegie classification and existence of a medical school within the institution. With a goal of attaining 18,000 eligible ECD respondents, the institutional level sample will include an initial sample of approximately 296 GSS institutions, 28 FFRDCs, and all 25 NIH IRP. All GSS institutions will be selected with probability proportionate to size (pps). The GSS pps size measure will be based on estimated overall number of ECD at each institution as well as the counts of ECD in subgroups defined by postdoc status, sex, race/ethnicity, and citizenship. The FFRDC will also be sampled using pps based on a simpler measure of size determined by the overall number of ECD and postdoc. All 25 NIH IRP will be sampled with certainty. At the second stage, samples of ECD stratified by postdoc status, sex, citizenship, and race/ethnicity will be selected within institutions.

Table 3 shows the 2017 ECDS anticipated first stage nonparticipation rates and second stage ineligibility and nonresponse rates. These estimates are based on the Pilot ECDS and assume slightly higher response rates due to the longer data collection period and changes to the ECD contacting protocols. Table 4 shows the 2017 ECDS institution and ECD estimated sample sizes.

Table 3. Anticipated Rates of Stage 1 Nonparticipation and Stage 2 Ineligibility and Nonresponse

	Stage 1 Sampling	Stage 2 Sampling	
Sampling stratum	% Not participating	% Ineligible	% Not responding[1]
GSS			
Medical schools and centers	15.0	3.0	25.0
Very high research activity universities	10.0	2.0	17.5
High research activity universities	15.0	1.0	17.5
All other colleges and universities	20.0	1.0	20.0
FFRDC	10.0	1.0	12.5
NIH IRP	0.0 [2]	1.0	30.0

[1] The "% Not responding" is among ineligible cases.

[2] The NIH Office of Intramural Training & Education will provide a list of ECDs for all 25 intramural research programs.

Table 4. ECDS Sampling Strata and Estimated Sample Sizes

	Institution Sample		ECD sample in institutions that provide lists	
Sampling stratum	Initial sample	Target completes (i.e., provide lists)	Initial sample	Target completes
GSS				
Medical schools and centers	63	53	5,136	3,736
Very high research activity universities	85	76	7,717	6,239
High research activity universities	64	54	3,682	3,007
All other colleges and universities	84	67	4,752	3,763
GSS total	296	250	21,283	16,745
FFRDC	28	25	976	845
NIH IRP	25	25	592	410
Total for all strata	349	300	22,855	18,000

The ECD sample is allocated to satisfy precision requirements set for the analytic domains. Analytic domains of interest include institution type, postdoc status, sex, U.S. citizenship, and race/ethnicity. Table 5 shows the minimum sample size and expected precision level by key analytic domains. Precision in this table is expressed as the coefficient of variation (CV) for an estimated proportion of 0.5 for the domain. Given the expected (or desired) CV, we calculated minimum sample size for each domain. This minimum sample size is then inflated to account for weight variation and survey nonresponse within domain. This inflated sample size becomes the initial sample size allocated to the domain. Attachment H gives further details of the sample design and the processes used to allocate the sample to the first and second stage strata.

Table 5. Minimum Sample Size and Coefficient Variation (CV) by Key Domains of Interest

Domain level	Category	Minimum sample size[1]	Expected CV[2]
Combined GSS, FFRDC, and NIH IRP Postdoc Status	Non-Postdoc	6,336	0.03
	Postdoc	5,979	0.03
GSS Institutions GSS, total		11,445	0.01
Sex	Female	4,237	0.02
	Male	4,259	0.02
Postdoc Status	Non-Postdoc	4,815	0.03
	Postdoc	4,588	0.03
Citizenship-Race-Ethnicity	Non-U.S. citizen	1,527	0.04
	U.S. citizen–White	5,981	0.02
	U.S. citizen–Asian	1,329	0.04
	U.S. citizen–Minority	1,622	0.04
GSS Substrata	Medical schools and centers	2,487	0.03
	Very high research activity universities	3,052	0.03
	High research activity universities	2,853	0.03
	All other colleges and universities	3,708	0.03
GSS substrata × Postdoc Status	Medical schools and centers; Postdoc	1,557	0.04
	Very high research activity universities; Postdoc.	2,508	0.04
	High research activity universities; Postdoc	383	0.10
	All other colleges and universities; Postdoc	270	0.10
	Medical schools and centers; Non-Postdoc	2,214	0.04
	Very high research activity universities; Non-Postdoc	2,253	0.04
	High research activity universities; Non-Postdoc	2,730	0.04
	All other colleges and universities; Non-Postdoc	3,267	0.03
Postdoc Status × Sex × Citizenship-Race-Ethnicity	Postdoc; Non-U.S. citizen; Female	921	0.04
	Postdoc; Non-U.S. citizen; Male	1,592	0.04
	Postdoc; U.S. citizen–White; Female	459	0.05
	Postdoc; U.S. citizen–White; Male	771	0.05
	Postdoc; U.S. citizen–Asian; Female	254	0.09
	Postdoc; U.S. citizen–Asian; Male	210	0.09
	Postdoc; U.S. citizen–Minority; Female	236	0.08
	Postdoc; U.S. citizen–Minority; Male	165	0.08
	Non-Postdoc; Non-U.S. citizen; Female	711	0.05
	Non-Postdoc; Non-U.S. citizen; Male	495	0.05
	Non-Postdoc; U.S. citizen–White; Female	999	0.05
	Non-Postdoc; U.S. citizen–White; Male	2,564	0.03
	Non-Postdoc; U.S. citizen–Asian; Female	1,071	0.05
	Non-Postdoc; U.S. citizen–Asian; Male	616	0.05
	Non-Postdoc; U.S. citizen–Minority; Female	765	0.05
	Non-Postdoc; U.S. citizen–Minority; Male	600	0.05
GSS substrata × Postdoc Status × Sex × Citizenship-Race-Ethnicity for first 2 GSS strata (Medical schools/	Med-schools; Postdoc; Non-U.S. citizen; Female	313	0.06
	Med-schools; Postdoc; Non-U.S. citizen; Male	505	0.06
	Med-schools; Postdoc; U.S. citizen–White; Female	222	0.08
	Med-schools; Postdoc; U.S. citizen–White; Male	223	0.08
	Med-schools; Postdoc; U.S. citizen–Asian; Female	109	0.10
	Med-schools; Postdoc; U.S. citizen–Asian; Male	110	0.10

centers, and Very high research activity only)[3]	Med-schools; Postdoc; U.S. citizen–Minority; Female	104	0.10
	Med-schools; Postdoc; U.S. citizen–Minority; Male	102	0.10
	Med-schools; Non-Postdoc; Non-U.S. citizen; Female	77	0.12
	Med-schools; Non-Postdoc; Non-U.S. citizen; Male	102	0.12
	Med-schools; Non-Postdoc; U.S. citizen–White; Female	429	0.08
	Med-schools; Non-Postdoc; U.S. citizen–White; Male	408	0.08
	Med-schools; Non-Postdoc; U.S. citizen–Asian; Female	178	0.08
	Med-schools; Non-Postdoc; U.S. citizen–Asian; Male	180	0.08
	Med-schools; Non-Postdoc; U.S. citizen–Minority; Female	143	0.10
	Med-schools; Non-Postdoc; U.S. citizen–Minority; Male	114	0.10
	Very-High-Research; Postdoc; Non-U.S. citizen; Female	341	0.07
	Very-High-Research; Postdoc; Non-U.S. citizen; Male	536	0.07
	Very-High-Research; Postdoc; U.S. citizen–White; Female	393	0.07
	Very-High-Research; Postdoc; U.S. citizen–White; Male	401	0.07
	Very-High-Research; Postdoc; U.S. citizen–Asian; Female	133	0.12
	Very-High-Research; Postdoc; U.S. citizen–Asian; Male	165	0.10
	Very-High-Research; Postdoc; U.S. citizen–Minority; Female	113	0.12
	Very-High-Research; Postdoc; U.S. citizen–Minority; Male	108	0.10
	Very-High-Research; Non-Postdoc; Non-U.S. citizen; Female	173	0.12
	Very-High-Research; Non-Postdoc; Non-U.S. citizen; Male	244	0.08
	Very-High-Research; Non-Postdoc; U.S. citizen–White; Female	1,067	0.05
	Very-High-Research; Non-Postdoc; U.S. citizen–White; Male	970	0.05
	Very-High-Research; Non-Postdoc; U.S. citizen–Asian; Female	239	0.12
	Very-High-Research; Non-Postdoc; U.S. citizen–Asian; Male	300	0.08
	Very-High-Research; Non-Postdoc; U.S. citizen–Minority; Female	178	0.12
	Very-High-Research; Non-Postdoc; U.S. citizen–Minority; Male	177	0.12
FFRDC Postdoc Status	Non-Postdoc	438	0.05
	Postdoc	406	0.05
NIH IRP Postdoc Status	Non-Postdoc	123	0.12
	Postdoc	287	0.08

[1] The counts in this column represent the minimum sample size needed to meet the domain level precision requirements specified by the CV in the next column. The minimum sample size was calculated using a formula that takes desired CV, proportion (we used a conservative 0.5) and design effect from the Pilot ECDS as the input.

[2] The expected (or desired) CVs are developed based on reviewing analytical goals and the estimated CVs achievable under the full sample size of 18,000.

[3] Constraints were not set for the High research activity and All other colleges and universities strata as the population sizes are so small in these domains that achieving adequate precision would require selecting a very high proportion of the ECD in these domains.

B.2   Description of Survey Methodology and Statistical Procedures

As noted above, the ECDS is a two-stage sample design, with an initial sampling at the institutional level and then a sampling of individual ECD from the ECD lists provided. Correspondingly, it is a two-stage data collection effort. The first stage of data collection involves contacting institutions to elicit their cooperation to compile a list of all ECD working at the institution. This process begins with the ECDS contractor identifying potential list coordinators (LCs) based on their GSS experience or position within the institution. With this information, a package will be sent to high authority figures (HAs)—university presidents at GSS institutions, the center director at FFRDCs, and the NIH IRP Director of the Office of Postdoctoral Services (OPS) in the Office of Intramural Training and Education (OITE)—seeking institutional participation. A follow-up call will be made to HAs seeking to clarify why participation is important and to further explain the steps required for the institution's participation.

Once the HA has approved the institution's participation in the ECDS and designated an LC, the LC is contacted to begin the process of compiling the list of ECD. Communications between the LC and ECDS contractor staff are on-going while the list is being compiled. Initial communications are made to establish a relationship and elicit cooperation. Follow-up communications are made to clarify the task requirements and prompt late responding LCs. Once the requested list of ECD and accompanying data are provided, the lists are reviewed for completeness and compared to the number of ECD expected based on institutional responses to the GSS, FFRDC Postdoc Survey, and IPEDS data collections, as well as institutional websites. ECDS contractor staff will follow-up with LCs until the list data are deemed final, and then the lists will be prepared for sampling and the second stage of data collection begins.

The second stage of data collection begins with a pre-notification sent by either the institutional HA or NSF/NCSES. For all institutions where HAs agree, pre-notification will be sent by the HA via e-mail, and the ECD's survey invitation will be sent via e-mail the following business day. However, it is known from the methodological study and Pilot ECDS that some HAs will decline to send the pre-notification or will be unable to do so within the necessary timeframe. In these cases, HAs will be asked to provide a letter of support to be sent with a hard-copy pre-notification letter from NSF/NCSES. For these cases that involve a pre-notification from NSF/NCSES, the invitation e-mail will be sent within five business days of pre-notification. Subsequent contacts are contingent upon completion status—i.e., follow-up contacts will cease once a survey is completed, it is determined from either the survey questions or communications from the prospective respondent that they are ineligible, or the sample member provides a firm refusal to participate.

Following Dillman's Tailored Design Method (TDM; Dillman 2007; Dillman et al. 2014)[1], second stage nonrespondents will receive a number of follow-up communications. In the order of their use, these include: two reminder e-mails, a reminder call, a third reminder e-mail, a mailed reminder letter with an accompanying brochure and letter of support if one was provided by the institutional HA, a second reminder call with a Computer-Assisted Telephone Instrument (CATI)

---

[1] Please see Attachment I for references.

option, and a final e-mail reminder. A final communication in the form of a thank you e-mail will be sent to all survey respondents.

A more detailed description of the contact strategies for both stages of data collection are provided in Attachment A. Copies of the contact materials can be found in Attachment B (institutional contacts) and Attachment C (ECD contacts).

B.2.1   Imputation for Item Nonresponse in the 2017 ECDS

Imputation is planned for all missing variables. The following variables are considered key and will be imputed first using frame information as well as external databases such as the Survey of Earned Doctorates (SED)[2]: field of study, origin (country) of degree, gender, citizenship, race/ethnicity. These variables are needed to establish national estimates of the total ECD population and key subpopulations. For the remaining survey items, items will be imputed in the order from least to greatest percentage of missing data, using a weighted hot deck procedure (Cox, 1980). For each imputed variable, an imputation class will be defined using variables that are previously imputed and highly correlate with the imputed variables. An imputation flag or indicator variable will be created and placed on the data file for all variables.

B.3   Methods Used to Maximize Response Rate

The following approaches will be used in the 2017 ECDS to maximize response rates, including features that were implemented successfully in the methodological study and Pilot ECDS as well as some additional first time approaches.

Stage 1: Institutional stage

- The initial request for participation is sent to institution HAs in a FedEx package and includes a letter and brochure that detail what participation will require. The package also includes a survey participation form that is pre-filled with the name of a potential LC.
- New for the 2017 ECDS, HAs will be asked to identify a communication coordinator (CC) to streamline the collection of letters of support and the sending of the pre-notification e-mails.
- LCs will receive an introductory call and information packet to confirm the appropriateness of their nomination and elicit cooperation.
- Follow-up e-mails and phone calls to LCs will offer help and seek updates on list compilation progress.
- New for the 2017 ECDS, a separate, secure section of the ECDS website for LCs that will provide answers to frequently asked questions, provide detailed instructions for preparing the ECD list, and enable secure upload of files.

Stage 2: ECD stage

---

[2] Missing frame variables are imputed during sample selection using multiple databases. Please see Attachment H for more details.

- A pre-notification message from the HA is provided to the ECD whenever possible. The type of pre-notification depends on the level of participation to which the HA agrees. Pre-notification can include either an e-mail directly from the HA or a letter of support the HA provides to the ECDS contractor that will be enclosed with a pre-notification letter from NSF/NCSES.

- There are several reminder e-mails to nonrespondents describing the importance of survey participation. All e-mails will include a hyperlink with embedded and secure login credentials making it easy for the prospective respondent to access the web survey.

- There are up to two reminder phone calls to nonrespondents aimed at ensuring that the ECD has received previous e-mail communications and to determine if the prospective respondent has encountered problems preventing them from completing the web survey. If the respondent prefers to complete the interview via phone, ECDS contractor staff will administer the interview over the phone.

- There is a hard-copy reminder mailing to nonrespondents. Two versions of the reminder letter will be used. The version sent to each ECD will be contingent upon whether pre-notification was sent by the HA or NSF/NCSES. When HAs have provided them, a second letter of support will be included in the reminder mailing.

In the first stage of data collection, 84.7% of the sampled institutions participated in the Pilot ECDS and provided lists of ECD working at the institution. A 66.3% response rate (AAPOR RR2[3]) was achieved at the second stage of data collection (e.g., among sample members) for the Pilot ECDS. For this response rate calculation, a partial interview was defined as any person who had not completed the full survey but had indicated in the first section of the survey that they had earned a doctorate or doctorate-equivalent degree within the previous 10 years and in the second section had indicated the types of work products they had produced. Due to lags in obtaining ECD lists from institutions and getting institutions to send the pre-notification e-mails, many sample members received an abbreviated version of the contacting protocol in the Pilot ECDS. Initially, the interval between reminders was shortened. Closer to the end of data collection, reminders were skipped. Response rates for sample members at institutions where nonrespondents received all of the reminders at the planned timing intervals was 79.4%.

With a few minor modifications, the methods to maximize response at the institutional level are the same or similar to those used in the Pilot ECDS. During stage one (institution level) contacting, the timing of a request for the letters of support has been moved back to provide time for HAs to become better acquainted with the survey before preparing the letter (see Attachment B). Additionally, communications with HAs and LCs have been revised to reflect a limited number of deadlines, which will facilitate increased use of automated batch e-mail communications. Another important addition for stage one is the section of the website that provides additional information to LCs. The resources on the web interface for the LCs are designed to address the most common questions that arose during the methodological study and Pilot ECDS.

---

[3] From AAPOR Standard Definitions, available for download at: http://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx

For the stage two (ECD level) data collection, a notable change to the contact strategy is to start the CATI phase with a prompting phone call, asking them to complete the survey through the Web instrument rather than CATI (see Attachment C). During the Pilot ECDS, the CATI outreach resulted in very few interviews and completing over the phone was significantly slower and more difficult than responding online. Therefore, the ECDS contractor staff will still administer the CATI interview to the ECD who prefer that method, but telephone interviewers will not encourage the use of CATI.

B.4   Testing of Procedures

The Pilot ECDS had several primary objectives: to create a methodology for building an ECD institutional sampling frame that could be used to develop national ECD estimates, and to test different strategies for contacting and recruiting sample members. This section provides an overview of the outcomes for those objectives. The section then discusses the goals and details associated with two methodological experiments proposed for inclusion in the 2017 ECDS. Finally, the section concludes with a discussion of the planned evaluation to determine the future of the ECDS and the role that the information collected in the 2017 ECDS survey will serve in the evaluation.

B.4.1   Test of Building Sample Frame

The first step in the Pilot ECDS was to further confirm the feasibility of building a sampling frame of ECD. The methodology for this step was modeled after the Occupational Information Network (O*NET), an establishment study for which the contractor has achieved good response rates over a period of ten years (e.g., establishment response rate of 76.1% and employee response rate of 65.0%). Protocols for the Pilot ECDS were designed to maximize institutional participation while minimizing the burden on institutional respondents. To achieve these two goals, the methodology leveraged contacts developed through the NSF-NIH GSS. When possible, the contractor identified a contact at institutions based on experience in the GSS and recommended that person to the HA as the person to perform the tasks necessary to compile a list of ECD at the institution (in the Pilot ECDS this person was referred to as the list coordinator [LC]). Follow-ups with the HA were made when necessary to seek institutional participation. Once a relationship was established with the LC, follow-up contacts were made to both assist the LC with any questions about the task and, when needed, prompt the LC about task completion.

Of the 176 unique institutions in the Pilot ECDS, 149 (84.7%) were able to compile and provide NCSES a list of ECD during the approximately 6 months when lists were accepted. As expected, this rate exceeded that from the methodological study, where 56 of the 81 sampled institutions (69.1%) provided an ECD list during a 3-month stage one data collection period.

Overall, 131 of the 149 institutions that participated in the Pilot ECDS (87.9%) provided final lists that covered all types of ECD, while the remaining 18 (12.1%) had some known limitation (e.g., missing medical residents). To assure such a high level of coverage, ECDS contractor staff conducted a more extensive review than the one used in the methodological study. The Pilot ECDS list review involved inspecting the lists LCs provided and comparing them to postdoc and nonfaculty researcher counts from the GSS, instructional staff data from IPEDS, FFRDC size

estimates, and the lists from the methodological study. When ECDS contractor staff identified a notable count deviation for a particular type of ECD or had reason to suspect that a category of employment had been erroneously included or excluded, they followed up with the LC to seek clarification.

While 87.9% of final lists from institutions in the Pilot ECDS included all potential ECD at the institution, the lists did not always have complete information for all individuals. For example, only 61.1% of lists had the doctorate type available for all individuals and only 47.0% of the lists had the doctorate year available for all individuals. For the potential ECD whose highest degree level or doctoral year was missing, NSF/NCSES requested listings for job titles that might include ECD. The ECDS contractor then estimated the likelihood that the person was an ECD based on available data (e.g., job title, age, postdoc indicator, hire year). Of the 6,827 sample members in the Pilot ECDS, 4,835 (70.8%) of the individuals on the sampling frame were expected to be ECD (i.e., administrative records indicated their first doctoral degree was awarded in 2004 or later), 1,221 (17.9%) were sampled as likely ECD, 639 (9.4%) were sampled as somewhat likely ECD, and 132 (1.9%) were sampled as unlikely ECD.[4]

B.4.2 Experimental Results for the Pilot ECDS

Two experiments were conducted for the Pilot ECDS. The first experiment assessed the effectiveness of different subject lines for the e-mail contacts. Some empirical evidence showed that the subject lines of e-mails can impact response rates (see Couper, 2008 for a summary). NSF/NCSES experimentally tested by increasing the sense of urgency in the subject line and the prominence of the survey sponsor. As shown in table 6, a control and two treatments were specified with experimental groups #1 and #2 providing greater detail and an increasing sense of urgency as compared to the control group. Experimental group #2 emphasized the National Science Foundation as the sponsor and deemphasized the survey title.

Table 6. E-mail Subject Line Experimental and Control Conditions

	Control	Experimental Group #1	Experimental Group #2
Login credentials E-mail	NSF Early Career Doctorates Survey	NSF Early Career Doctorates Survey – Your Login Credentials	National Science Foundation ECD Survey – Your Login Credentials
Reminder E-mail #1	NSF Early Career Doctorates Survey	NSF Early Career Doctorates Survey – Reminder	National Science Foundation ECD Survey – Reminder
Reminder E-mail #2	NSF Early Career Doctorates Survey	NSF Early Career Doctorates Survey – Your Help Needed	National Science Foundation ECD Survey – Your Help Needed

---

[4] The Pilot ECDS ineligibility rate of respondents by ECD sampling strata (expected, likely, somewhat likely, and unlikely) were consistent with their final disposition: 3.3% of expected ECD were found to be ineligible (e.g., individual had earned a first doctoral degree prior to 2004 or were no longer at the sampled institution), 7.7% of likely ECD were known to be ineligible, 31.1% of somewhat likely ECD were known to be ineligible, and 53.8% of unlikely ECD were known to be ineligible.

| Reminder E-mail #3 | NSF Early Career Doctorates Survey | NSF Early Career Doctorates Survey – Please Respond | National Science Foundation ECD Survey – Please Respond |

Results presented in Table 7 suggest that the subject line had no significant impact on response among eligible respondents who were invited to participate.[5] In the control group and the two experimental groups, 69% or 70% of ECD responded.

Table 7. Response by E-mail Subject Line Treatment

Treatment	Number of Respondents	Response Rate
Total	4,179	69.6
Control Group	1,398	70.2
Experimental Group 1	1,398	69.6
Experimental Group 2	1,383	69.1

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Early Career Doctorates Survey (ECDS) pilot study, 2015.

A second experiment in the Pilot ECDS tested the hypothesis that a priority mailing would improve response rates more than a regular United States Postal Service (USPS) first class mailing. In the Pilot ECDS, the sixth contact was a mailing to all non-responding ECD. This mailing included a cover letter, brochure and a letter of support from the HA, when applicable. According to Dillman's TDM approach (2014), letters sent via priority mail differ in the packaging, mode of delivery, and speed with which they are delivered. This makes the letter more noticeable to the recipient before they open the package and is expected to have the effect of lending greater importance and legitimacy of the survey request. If regular mail results in comparable response rates to the priority mail method, substantial savings would be gained in future waves of the survey.

For this experiment, sample members were divided into two treatment groups. One group received the reminder letter via regular USPS first class mail while the other received a priority FedEx or USPS Priority Mail package. Overall, as shown in Table 8, the response rate for sample members who received the priority mailing was about 8 percentage points higher than the response rate for those who received the USPS first class mailing (49.9 vs. 41.1 percent; $p < .01$). Based on this result, the 2017 ECDS will use USPS Priority Mail for this contact.

Table 8. Response by Mail Class Treatment

Treatment	Number of Respondents	Response Rate
Total	478	45.5
USPS First Class Mail	217	41.1
FedEx / USPS Priority Mail	261	49.9

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Early Career Doctorates Survey (ECDS) pilot study, 2015.

---

[5] Sample members who opted out or did not opt in did not receive any e-mails and are therefore excluded from this analysis.

B.4.3 Experiments for the 2017 ECDS

Two experiments are proposed for the 2017 ECDS. The first will evaluate delivery methods for the pre-notification contact in the NSF/NCSES protocol group. In the past, sample members in the NSF/NCSES protocol group received a pre-notification from NSF/NCSES in the form of a hard copy letter sent by regular USPS first class mail. In the 2017 ECDS, an experiment will test the effect of various modes of delivery of this NSF/NCSES pre-notification on response rates and survey costs. In the 2017 ECDS, ECD in the NSF/NCSES protocol group may receive the pre-notification in one of three ways depending on experimental group; regular USPS first class mail, priority mail (i.e., FedEx or priority U.S. mail), or e-mail. We expect that 15-20% of institutions will not send the pre-notification e-mails, resulting in 3,000 to 4,000 sample members being placed in this experiment. Systematic assignment to groups will ensure even allocation across the three treatments.

There are advantages to e-mail over USPS first class mail in terms of speed of delivery, predictability of delivery timing, nearly immediate indication of failed delivery, and cost. On the other hand, a hard copy letter's formality may carry more weight with prospective respondents than e-mail. Based on the results of the reminder letter experiment conducted in the Pilot ECDS (see B.4.2), it is expected that sending hard copy pre-notifications by priority mail will yield higher response rates than sending by regular USPS first class mail. The special packaging and hastened delivery of priority mail may convey importance and have a positive effect on response rates. In addition, priority mail has practical advantages over USPS first class mail. With priority mail the ECDS contractor knows the date of delivery and is informed if the mailing cannot be delivered. However, priority mail is more expensive than regular USPS first class mail. Comparing the effectiveness and cost of each of these three modes of delivery in an experiment will help NSF/NCSES choose the approach that maximizes response rates while balancing costs in future ECDS data collections.

A second experiment is proposed to evaluate the feasibility of collecting some of the information asked in the ECDS questionnaire from other sources, the reliability of these data, and the reduction in burden that could be achieved if ECD were no longer asked to provide this information. Two alternative sources of information will be investigated. The first source is ECD's curriculum vitae. The other source is an online author identifier (e.g., ORCID, Researcher ID, Scopus) that provides a unique identifier to registrants and stores information related to their education and research activities. NCSES estimates that the data gleaned from these sources has the potential to eliminate up to 6 minutes from the survey instrument, and will cover the following topics: educational history (types of postsecondary degrees earned, degree granting institutions, dates of award, and fields of study), employment history (employer names, job titles, and dates of employment), and work product information (presentations and publications).

We anticipate that there may be differences in organization and content of a curriculum vitae as well as usage of online author identifiers by country of citizenship and origin of degree. Therefore, we propose two basic experimental groups – curriculum vitae and online author identifier – of 300 responding ECD, each made up of 100 responding ECD from the following three substrata: U.S. citizens, Non-U.S. citizens with doctorate degrees from U.S. institutions, and Non-U.S. citizens with doctorate degrees from non-U.S. institutions. A sequential

assignment routine (i.e., every Nth respondent with the required characteristics) will be built into the questionnaire with rates differing by expected proportions such that equal numbers of ECD will be sampled within each group and strata.

At the end of the questionnaire, selected ECD will be informed of the experiment, asked to upload their curriculum vitae or enter their online author identifier, and provided the option to decline. ECD in the online author identifier group will be given an option to indicate that they do not know what an online author identifier is or that they know what an online author identifier is but do not have an identifier. ECD in the curriculum vitae group will be able to upload it in any format (e.g., pdf, Microsoft Word document).

Comparisons will include the rate at which ECD agree to provide their curriculum vitae or an online author identifier, the rates at which specified data elements are found on the curriculum vitae or the online author identifier profile, and rates of agreement of those data elements with data provided in the survey instrument. These rates will be compared for curriculum vitaes and online author identifier profiles and for each of the three sampling strata.

B.4.4 Changes to the ECDS Questionnaire

The ECDS questionnaire (Attachment E) was revised following the Pilot ECDS to improve problematic items, add options to items where "other specify" responses indicated common responses not already listed, and to make the survey more applicable to ECD in clinical fields and the humanities. NCSES held a Human Resources Expert Panel (HREP) meeting to review the questionnaire and help the project staff identify areas where the questionnaire could be cut. In the Pilot ECDS, the time to complete the survey was approximately 41 minutes. Based on changes to the survey suggested by the HREP and analysis of the Pilot ECDS response data, NCSES estimates the completion time for the 2017 ECDS questionnaire will be approximately 32 minutes.

Major changes to the questionnaire include:

- Removing or streamlining questionnaire response options based on the Pilot ECDS results to improve item salience and to reduce time needed to complete the survey[6]
- Asking for the specific doctoral degree (e.g., PhD, MD, EdD, etc.) and using that response as a fill throughout the survey instrument (as opposed to asking whether it was a doctorate or doctorate-equivalent first and then using the terms doctorate or doctorate-equivalent throughout)
- Changing the order of the employment loop (section C) to ask about the current or most recent job at the sampled institution first, followed by the first position, and then a postdoc if the ECD had indicated having a postdoc and they had not reported on it in the first two iterations

---

[6] Based on the Pilot ECDS results, NCSES deleted response options from questionnaire items if they were not frequently used and added response options to questionnaire items based on information provided in the "other specify" category.

- Removing the full-time, paid requirements for the first position to capture the bridge positions that many ECD take immediately after earning their degree
- Revising and reordering complex items taken from the OECD Survey of Career Doctorate Holders to reduce the cognitive effort and time needed to answer these questions[7]

B.4.5 Evaluation to Determine the Future of the Early Career Doctorates Project

The Early Career Doctorates Project was established by NCSES to gather in-depth information about ECD, including postdocs. Through the work that began with the multi-year Postdoc Data Project, continued with the Pilot ECDS, and will reach full-scale survey production with the 2017 ECDS, NCSES has made strides in addressing the coverage and data limitations issues that have historically plagued the attempt to better understand the ECD population. As with prior ECD efforts, at the conclusion of the 2017 ECDS, NCSES plans to conduct an evaluation of the results to determine whether the Early Career Doctorates Project is actually filling the gaps as designed in NCSES's coverage and content related to the doctorate population.

The goal for this evaluation is to identify the role of the ECDS in the context of NCSES's suite of science and engineering (S&E) workforce surveys. Specifically, the ECDS was designed to address the coverage issues and lack of data on work experiences that has limited the understanding of the ECD population. However, there is overlap in the target population and survey content for the ECDS with other NCSES S&E workforce surveys including the Survey of Doctorate Recipients (SDR), Survey of Earned Doctorates (SED), and National Survey of College Graduates (NSCG). This evaluation should assess the unique purpose, population coverage, and data utility for each survey, and provide insight on how the suite of S&E workforce surveys jointly support NCSES's role in the collection, interpretation, analysis, and dissemination of objective data on the S&E enterprise. The findings from this evaluation will guide the future decisions for the ECDS including the timing, population coverage, survey content, and sample design for any future surveys of the ECD population.

NCSES plans to use the following sources of information to aid in the evaluation:

- A post-survey assessment of the population coverage and estimation reliability of the 2017 ECDS data, the long-term feasibility of collecting ECD data, and the utility of the ECDS data products. This assessment will include a comparison of ECD estimates from the 2017 ECDS with ECD estimates from other NCSES surveys. This comparison will provide insight on the differences across the NCSES surveys in the coverage and reliability for estimates of the ECD population. As an example, to assess the coverage and reliability of the 2017 ECDS estimates of postdocs, NCSES will compare the 2017 ECDS estimates for postdocs in three employment settings (U.S. academic institutions,

---

[7] In the Pilot ECDS, it took an average of 3.6 minutes for ECD to respond to items A12 (preparation for employment by graduate program), E1 (meeting supervisor's expectations), and E2 (perceived value of degree). By revising the response options, placing A12 immediately after E1, and conditioning the subitems in A12 on the responses to corresponding sub-items within E1, we estimate that response time across these three items will be reduced to just over 2 minutes.

FFRDCs, and the NIH IRP) against postdoc estimates from the current NCSES surveys (e.g., SDR, SED, GSS, and the Survey of Postdocs at FFRDCs).

- Follow-up discussions with key ECDS stakeholders including, but not limited to, the National Institutes of Health, the National Postdoctoral Association, and the American Association of Medical Colleges. These discussions will attempt to better understand the current alignment between the ECDS data and the stakeholders' data needs.
- Findings from the National Academies of Science, Engineering, and Medicine's Committee on National Statistics panel examining the NCSES effort to measure the S&E workforce. This panel will issue a report in late 2017 with recommendations for improving the relevance, accuracy, timeliness, and cost-effectiveness of S&E workforce data. The information included in this report will provide details, direction, and guidance necessary for NCSES to develop a robust and flexible framework for measuring the S&E workforce over the coming decades.

B.5   Names and Telephone Numbers of Individuals Consulted

The individuals consulted on the ECDS technical and statistical issues are listed in table 9.

Table 9. Individuals Consulted on ECDS Technical and Statistical Issues

Name	Affiliation	Telephone number
Ms. Kelly Phou ECDS Manager	National Science Foundation, NCSES Arlington, VA	703-292-7422
Mr. John Finamore Program Director	National Science Foundation, NCSES Arlington, VA	703-292-2258
Ms. Emilda Rivers Deputy Director	National Science Foundation, NCSES Arlington, VA	703-292-7773
Mr. John R. Gawalt Director	National Science Foundation, NCSES Arlington, VA	703-292-7776
Ms. Jeri Mulrow Former Deputy Director	National Science Foundation, NCSES Arlington, VA	
Dr. Samson Adeshiyan Chief Statistician	National Science Foundation, NCSES Arlington, VA	703-292-7769
Dr. Stephen Cohen Former Chief Statistician	National Science Foundation, NCSES Arlington, VA	
Dr. Wan-Ying Chang Mathematical Statistician	National Science Foundation, NCSES Arlington, VA	703-292-2310
Mr. Darius Singpurwalla Mathematical Statistician	National Science Foundation, NCSES Arlington, VA	703-292-7793
Ms. Rebecca Morrison	National Science Foundation, NCSES	703-292-7794

Survey Methodologist	Arlington, VA	
Ms. Jennifer Sutton Research Training Coordinator	National Institutes of Health Bethesda, MD	301-435-2686
Dr. Lori Conlan Director, Career Services Center	National Institutes of Health Bethesda, MD	301-435-7231
Ms. Cathee Johnson Phillips Former Executive Director	National Postdoctoral Association Washington, DC	
Dr. Jodi Yellin Director, Science Policy	American Association of Medical Colleges, Washington, DC	202-828-0485
Dr. Irena Tartatkovsky Senior Science Policy Analyst	American Association of Medical Colleges, Washington, DC	202-862-6134
Ms. Roxanne Murray Former Director, HR & Administration	Association of American Universities Washington, DC	
Mr. Peter Einaudi Project Director	RTI International Research Triangle Park, NC	919-541-8765
Dr. Paul Biemer Senior Survey Methodologist	RTI International Research Triangle Park, NC	919-541-6056
Ms. Laura Burns Fritch Survey Methodologist	RTI International Research Triangle Park, NC	919-990-8318
Dr. Sara Wheeless Mathematical Statistical Task Leader	RTI International Research Triangle Park, NC	919-541-5891
Dr. Amang Sukasih Senior Statistician	RTI International Washington, DC	202-728-2469
Dr. Patricia Green Senior Advisor	RTI International Chicago, IL	312-456-5260
Mr. Bob Steele Systems Development Task Leader	RTI International Research Triangle Park, NC	919-316-3836
Dr. Caren Arbeit Senior Analyst	RTI International Berkeley, CA	510-818-4812