

**SF-83-1 SUPPORTING STATEMENT:
SECTION B**

for the

2019

National Survey of College Graduates

Contents

B.	COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS.....	1
1.	RESPONDENT UNIVERSE AND SAMPLING METHODS.....	1
2.	SURVEY METHODOLOGY.....	3
3.	METHODS TO MAXIMIZE RESPONSE.....	9
4.	TESTING OF PROCEDURES.....	11
5.	CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION.....	18

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

When the American Community Survey (ACS) replaced the decennial census long form beginning with the 2010 Census, the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) identified the ACS as the potential sampling frame for the National Survey of College Graduates (NSCG) for use in the 2010 survey cycle and beyond. After reviewing numerous sample design options proposed by NCSES, the Committee on National Statistics (CNSTAT) recommended a rotating panel design for the 2010 decade of the NSCG (National Research Council, 2008). In this design, a new ACS-based sample of college graduates will be selected and followed for four biennial cycles before the panel is rotated out of the survey. The use of the ACS as a sampling frame, including the field of degree questionnaire item included on the ACS, allows NCSES to more efficiently target the science and engineering (S&E) workforce population. Furthermore, the rotating panel design planned for the 2010 decade allows the NSCG to address certain deficiencies¹ of the previous design, including the undercoverage of key groups of interest such as foreign-degreed immigrants with S&E degrees and individuals with non-S&E degrees who are working in S&E occupations.

The NSCG design for the 2010 decade oversamples cases in small cells of particular interest to analysts, including underrepresented minorities, persons with disabilities, and non-U.S. citizens. The goal of this oversampling effort is to provide adequate sample for NSF's congressionally mandated report on *Women, Minorities, and Persons with Disabilities in Science and Engineering*.

The 2017 survey cycle marked the full implementation of the four-panel rotating panel design that began with the 2010 NSCG. Under this fully implemented rotating panel design, the 2019 NSCG will include 148,000 sample cases which includes:

- 1) Returning sample from the 2017 NSCG (originally selected from the 2011 ACS);
- 2) Returning sample from the 2017 NSCG (originally selected from the 2013 ACS);
- 3) Returning sample from the 2017 NSCG (originally selected from the 2015 ACS); and
- 4) New sample selected from the 2017 ACS.

About 67,000 new sample cases will be selected from the 2017 ACS. This sample size includes an increase of approximately 19,000 cases compared to the 2017 NSCG new sample to address three issues related to the NSCG design:

- (1) Unit nonresponse in subsequent survey cycles has been increasing making it more difficult for the NSCG to achieve its target reliability requirements for key analytical domains. Consequently, the size of the new sample will be increased by about 14,000 cases to offset the impact of future nonresponse among the returning sample and enable the NSCG to meet its target reliability requirements described in Appendix D.

¹ Prior to 2010, the NSCG selected its sample once each decade from the decennial census long form. NSCG respondents educated or working in S&E fields were then followed biennially throughout the decade. Since no additional NSCG sample was selected throughout the decade, the previous NSCG design suffered from undercoverage of immigrants who entered the U.S. during the decade and individuals who began working in an S&E occupation during the decade.

(2) Weight variation that exists due to the sampling rate differences of S&E and non-S&E occupations across sampling strata results in reduced precision of the estimates when cases sampled as non-S&E turn out to be S&E. To decrease this impact, the sample size for non-S&E cases in the new sample will be increased by 2,500 to 5,000. Similar to the 14,000-case increase described above, this sample increase for non-S&E cases will enable the NSCG to better meet its target reliability requirements described in Appendix D.

(3) Interest in individuals with foreign-earned doctorates has been increasing. To improve the NSCG’s estimation capability for foreign-earned doctorate recipients, a CNSTAT panel recently recommended that NCSES evaluate whether modeling techniques could be used to more efficiently identify these individuals.² To have sufficient data to conduct such an evaluation, 3,000 additional cases with doctorate degrees and a low likelihood of having earned the degree in the U.S. will be selected as part of the new sample.

The remaining 81,000 cases will be selected from the set of returning sample members. While most of the returning sample cases are respondents from the 2017 NSCG survey cycle, about 13,000 nonrespondents from the 2017 NSCG survey cycle will be included in the 2019 NSCG sample. These 13,000 cases are individuals who responded in their initial survey cycle but did not respond during the 2017 NSCG survey cycle. As was the case in previous NSCG years, previous-cycle nonrespondents are being included in the 2019 NSCG sample in an effort to reduce the potential for nonresponse bias in our NSCG survey estimates.

The table below provides an overview of the NSCG sample sizes throughout the 2010 decade.

NSCG Year	Panel				
2010	Returning Sample		New Sample		2010 Total
	2008 NSCG & RCG (n=35,000)		2009 ACS (n=65,000)		100,000
2013	Returning Sample		New Sample		2013 Total
	2010 RCG (n=13,000)	2009 ACS (n=47,000)	2011 ACS (n=83,000)*		143,000
2015	Returning Sample			New Sample	2015 Total
	2010 RCG (n=12,000)	2009 ACS (n=23,000)	2011 ACS (n=60,000)	2013 ACS (n=40,000)*	135,000
2017	Returning Sample			New Sample	2017 Total
	2009 ACS (n=22,500)	2011 ACS (n=28,000)	2013 ACS (n=25,000)	2015 ACS (n=48,500)*	124,000
2019	Returning Sample			New Sample	2019 Total
	2011 ACS (n=27,500)	2013 ACS (n=24,500)	2015 ACS (n=29,000)	2017 ACS (n=67,000)*	148,000

NOTES: * indicates the inclusion of a young graduates oversample
RCG refers to the National Survey of Recent College Graduates

² <https://www.nap.edu/catalog/24968/measuring-the-21st-century-science-and-engineering-workforce-population-evolving> (Recommendation 4.2 on p. 77).

The 2019 NSCG survey target population includes all U.S. residents under age 76 with at least a bachelor's degree prior to January 1, 2018. The new sample portion of the 2019 NSCG will provide complete coverage of this target population. The returning sample, on the other hand, will provide only partial coverage of the 2019 NSCG target population. Specifically, the returning sample will cover the population of U.S. residents under age 76 with at least a bachelor's degree prior to January 1, 2016. The returning sample also does not include recent immigrants who have come to the country since 2015.

There are several advantages of this rotating panel sample design. It: 1) permits benchmarking of estimates to population totals derived from the sample using the ACS; 2) maintains the sample sizes of and allows cross-sectional estimation for small populations of scientists and engineers of great interest such as underrepresented minorities, persons with disabilities, and non-U.S. citizens; 3) provides an oversample of young graduates to allow continued detailed estimation of the recent college graduates population; and 4) permits longitudinal analysis of the retained cases from the ACS-based sample.

Using the 2017 NSCG final response rates as a basis, NCSES estimates the final response rate for the 2019 NSCG to be 70 to 80 percent.

2. SURVEY METHODOLOGY

Sample Design and Selection

As part of the 2019 NSCG sample selection, the returning sample portion of the NSCG sampling frame will be sampled separately from the new sample portion.

The 2019 NSCG returning sample will be selected with certainty from the eligible cases on the returning sampling frame.

The sample selection for the 2019 NSCG new sample will use stratification variables similar to what was used in the 2017 NSCG. These stratification variables will be formed using response information from the 2017 ACS. The levels of the 2019 NSCG new sample stratification variables are as follows:

Highest Degree Level

- Bachelor's degree or professional degree
- Master's degree
- Doctorate degree

Occupation/Degree Field

A composite variable composed of occupation and bachelor's degree field of study

- Mathematicians
- Computer Scientists
- Biological/Medical Scientists
- Agricultural and Other Life Scientists
- Chemists
- Physicists and Other Physical Scientists
- Economists
- Social Scientists
- Psychologists

- Chemical Engineers
- Civil and Architectural Engineers
- Electrical Engineers
- Mechanical Engineers
- Other Engineers
- Health-related Occupations
- S&E-Related Non-Health Occupations
- Postsecondary Teacher, S&E Field of Degree
- Postsecondary Teacher, Non-S&E Field of Degree
- Secondary Teacher, S&E Field of Degree
- Secondary Teacher, Non-S&E Field of Degree
- Non-S&E High Interest Occupation, S&E Field of Degree
- Non-S&E Low Interest Occupation, S&E Field of Degree
- Non-S&E Occupation, Non-S&E Field of Degree
- Not Working, S&E Field of Degree or S&E Previous Occupation (if previously worked)
- Not Working, Non-S&E Field of Degree and Non-S&E Previous Occupation (if previously worked) or never worked

Demographic Group

A composite demographic variable composed of race, ethnicity, disability status, citizenship, and U.S.-earned degree status

- U.S. Citizen at Birth (USCAB) or non-USCAB with high likelihood of U.S.-earned degree, Hispanic
- USCAB or non-USCAB with high likelihood of U.S.-earned degree, Black
- USCAB or non-USCAB with high likelihood of U.S.-earned degree, Asian
- USCAB or non-USCAB with high likelihood of U.S.-earned degree, AIAN or NHPI³
- USCAB or non-USCAB with high likelihood of U.S.-earned degree, disabled
- USCAB or non-USCAB with high likelihood of U.S.-earned degree, White or Other
- Non-USCAB with low likelihood of U.S.-earned degree, Hispanic
- Non-USCAB with low likelihood of U.S.-earned degree, Asian
- Non-USCAB with low likelihood of U.S.-earned degree, remaining cases

In addition, for the sampling cells where a young graduate oversample is desired,⁴ an additional sampling stratification variable will be used to identify the oversampling areas of interest. The following criteria define the cases eligible for the young graduate oversample within the 2019 NSCG.

- 2017 ACS sample cases with a bachelor's degree who are age 30 or less and are educated or employed in an S&E field
- 2017 ACS sample cases with a master's degree who are age 34 or less and are educated or employed in an S&E field

The multiway cross-classification of these stratification variables produces approximately 1,000 non-empty sampling cells. This design ensures that the cells needed to produce the small demographic/degree field groups for the congressionally mandated report on *Women, Minorities and Persons with Disabilities in Science and Engineering* (see 42 U.S.C. 1885d) will be maintained.

³ AIAN = American Indian / Alaska Native, NHPI = Native Hawaiian / Pacific Islander

⁴ Since the young graduate oversample planned for the NSCG serves to offset the discontinuation of the National Survey of Recent College Graduates (NSRCG), the oversample will focus only on bachelor's and master's degree recipients as had the NSRCG.

The 2019 NSCG reliability targets are aligned with the data needs for the NSF congressionally mandated reports. The sample allocation will be determined based on reliability requirements for key NSCG analytical domains provided by NCSES. The 2019 NSCG coefficient of variation targets that drive the 2019 NSCG sample allocation and selection are included in Appendix D. Tables 1, 2, and 3 of Appendix D provide reliability requirements for estimates of the total college graduate population. Tables 4, 5, and 6 of Appendix D provide reliability requirements for estimates of young graduates, which are the target of the 2019 NSCG oversampling strata.

In total, the ACS-based sampling frame for the 2019 NSCG new sample portion includes approximately 1,016,000 cases representing the college-educated population of 69 million residing in the U.S. as of 2017. From this sampling frame, 67,000 new sample cases will be selected based on the sample allocation reliability requirements discussed in the previous paragraph.

Weighting Procedures

Estimates from the 2019 NSCG will be based on standard weighting procedures. As was the case with sample selection, the weighting adjustments will be done separately for the new sample cases and separately for each panel within the returning sample cases. The goal of the separate weighting processes is to produce final weights for each panel that reflect each panel's respective population. To produce the final weights, each case will start with a base weight defined as the probability of selection into the 2019 NSCG sample. This base weight reflects the differential sampling across strata. Base weights will then be adjusted to account for unit nonresponse.

Weighting Adjustment for Survey Nonresponse

Following the weighting methodology used in the 2017 NSCG, we will use propensity modeling to account and adjust for unit nonresponse. Propensity modeling uses logistic regression to determine if characteristics available for all sample cases, such as prior survey responses and paradata, can be used to predict response. One advantage to this approach over the cell collapsing approach used in the 1990 and 2000 decades of the NSCG is the potential to more accurately reallocate weight from nonrespondents to respondents that are similar to them, in an attempt to reduce nonresponse bias. An additional advantage to using propensity modeling is the avoidance of creating complex noninterview cell collapsing rules.

We will create a model to predict response using the sampling frame variables that exist for both respondents and nonrespondents. A logistic regression model will use response as the dependent variable. The propensities output from the model will be used to categorize cases into cells of approximately equal size, with similar response propensities in each cell. The noninterview weighting adjustment factors will be calculated within each of the cells.

The noninterview weighting adjustment factor is used to account for the weight of the 2019 NSCG nonrespondents when forming survey estimates. The weight of the nonrespondents will be redistributed to the respondents and ineligibles within the 2019 NSCG sample. After the noninterview adjustment, weights will be controlled to ACS population totals through a post-stratification procedure that ensures the population totals are upheld.

Weighting Adjustment for Extreme Weights

After the completion of these weighting steps, some of the weights may be relatively large compared to other weights in the same analytical domain. Since extreme weights can greatly increase the variance of survey estimates, NCSES will implement weight trimming options. When

weight trimming is used, the final survey estimates may be biased. However, by trimming the extreme weights, the assumption is that the decrease in variance will offset the associated increase in bias so that the final survey estimates have a smaller mean square error. Depending on the weighting truncation adjustment used to address extreme weights, it is possible the weighted totals for the key marginals will no longer equal the population totals used in the iterative raking procedure. To correct this possible inequality, the next step in the 2019 NSCG weighting processing will be an iterative raking procedure to control to pre-trimmed totals within key domains. Finally, an additional execution of the post-stratification procedure to control to ACS population totals will be performed.

Degree Undercoverage Adjustment

The 2019 NSCG new sample does not have complete coverage of the population that first earned a degree during 2017, the ACS data collection year. For example, an ACS sample person that earned their first degree in May 2017, would be eligible for selection into the NSCG if their household was interviewed by ACS in July 2017 (i.e., after they earned their first degree). However, they would not be eligible for selection into the NSCG if their household was interviewed by ACS in March 2017 (i.e., before they earned their first degree).

Given that individuals who earned a degree after their ACS interview date are not eligible for the NSCG, the 2019 NSCG has undercoverage of individuals with their first degrees earned in 2017. To ensure the 2019 NSCG provides coverage of all individuals with degrees earned during 2017, a weighting adjustment is included in the 2019 NSCG weighting processing to account for this undercoverage. The Census Bureau conducted research on weighting adjustment methods during the 2013 NSCG cycle and an iterative reweighting model-based method was chosen to adjust the weights for this undercoverage. The weights after the degree undercoverage adjustment serve as the final panel-level weights.

Derivation of Combined Weights

To increase the reliability of estimates of the small demographic/degree field groups used in the congressionally mandated report on *Women, Minorities and Persons with Disabilities in Science and Engineering* (see 42 U.S.C. 1885d), NCSES will combine the new sample and returning sample together and will form combined weights to use in estimation for the combined set of cases. The combined weights will be formed by adjusting the new sample final weights and the returning sample final weights to account for the overlap in target population coverage. The result will be a combined final weight for all 148,000 NSCG sample cases.

Replicate Weights

Sets of replicate weights will also be constructed to allow for separate variance estimation for the new sample and for each panel within the returning sample. The replicate weight for the combined estimates will be constructed from these sets of replicate weights. The entire weighting process applied to the full sample will be applied separately to each of the replicates in producing the replicate weights.

Standard Errors

The replicate weights will be used to estimate the standard errors of the 2019 NSCG estimates. The variance of a survey estimate based on any probability sample may be estimated by the method of replication. This method requires that the sample selection and the estimation procedures be independently carried through (replicated) several times. The dispersion of the resulting replicated estimates then can be used to measure the variance of the full sample.

Questionnaires and Survey Content

As was the case in the 2017 NSCG, we will use three different versions of the 2019 NSCG questionnaire: 1) one for new sample members, 2) one for returning sample members who responded in 2017, and 3) one for returning sample members who did not respond in 2017. The main difference is that the questionnaires for returning sample members do not include questions where the response likely would not change from one cycle to the next. Specifically, the questionnaire for new sample members includes a degree history grid and certain demographic questions (e.g., race, ethnicity, and gender) that are not asked in the questionnaires for the returning sample members. If these items were not reported by the returning sample members during a prior NSCG data collection, the web and CATI instruments will attempt to collect the information this cycle. The two questionnaires for the returning sample members are similar to one another, with the exception of a slightly longer date range for the questions about recent educational experiences for the previous cycle nonrespondents.

The survey questionnaire items for the NSCG can be divided into two types of questions: core and module. Core questions are defined as those considered to be the base for the surveys, including job characteristics, education activities, and demographics. These items are essential for sampling, respondent verification, basic labor force information, and/or robust analyses of the S&E workforce. They are asked of all respondents each time they are surveyed, as appropriate, to establish the baseline data and to update the respondents' labor force status and changes in employment and other demographic characteristics. Module items are defined as special topics that are asked less frequently on a rotational basis of the entire target population or some subset thereof. Module items tend to provide the data needed to satisfy specific policy, research, or data user needs. Past NSCG questionnaire modules have included federal support of work, job satisfaction, and immigration information.

Appendix E includes the 2019 NSCG questionnaire for the new sample members. The other two NSCG questionnaires (questionnaire for previous cycle respondents and the questionnaire for previous cycle nonrespondents) both include a subset of the questions included on the new sample questionnaire.

Non-Sampling Error Evaluation

In an effort to account for all sources of error in the 2019 NSCG survey cycle, the Census Bureau will produce a report that will include information similar in content to the 2017 NSCG Non-Sampling Error Report.⁵ The 2019 NSCG Non-Sampling Error Report will evaluate two areas of non-sampling error – nonresponse error and error as a result of the inconsistency between the ACS and NSCG responses. These topics will provide information about potential sources of non-sampling error for the 2019 NSCG survey cycle.

Nonresponse Error

Numerous metrics will be computed to motivate a discussion of nonresponse – unit response rates, compound response rates, estimates of key domains, and R-indicators. Each of these metrics provides different insights into the issue of nonresponse and will be discussed individually and then summarized together.

Unit response rates are a simple method of quantifying what percentage of the sample population responded to the survey. For example, in the 2017 NSCG new sample portion (sampled out of the

⁵ Creamer, Selmin “2017 NSCG Non-Sampling Error Report,” Census Bureau Memorandum from Tersine to Milan, July 2018, draft.

2015 ACS), the overall weighted response rate was 65.4%; however, individual demographic subgroups (e.g., Black, Hispanic, etc.) have weighted response rates that vary between 49.5% and 69.4%. Age groups had weighted response rates ranging from 52.8% for younger age groups to 71.5% for the oldest age groups. Some variation in response is expected due to random variation; however, large variations in response behavior can be a cause for concern with the potential to introduce nonresponse bias. Assuming we are measuring different subgroups of the target population separately because we are interested in the different response data they provide, having differential response rates across subgroups may mean we are missing information in the less responsive subgroups. This is the driving force behind nonresponse bias – a relationship between the explanatory variables and the outcome variables. If the explanatory variables are also related to the likelihood to respond, resulting estimates may be biased.

The compound response rate looks at response rates over time and considers how attrition can affect the respondent population. Attrition is important when considering the effect of nonresponse in rotating panel surveys like the NSCG. As an example, for the 2010 NSCG (sampled out of the 2009 ACS), a response rate of 98% in the ACS followed by a response rate of 73.4% in the 2010 NSCG and 64.5% in the 2017 NSCG results in a compound response rate of just 46.4%. This means that only 46% of the cases originally eligible and sampled for the ACS have responded in the 2017 NSCG. Attrition can lead to biased estimates, particularly for surveys that do not continue to follow nonrespondents in later rounds. This is because weighting adjustments and estimates are made on a dwindling portion of the population. This can lead to weight inflation and increased variances, which may make significant differences more difficult to detect in the population. Further, if respondents are different (e.g., would provide different information) from nonrespondents, excluding the nonrespondents effectively excludes a portion of valuable information from the response and the resulting estimates. The estimates become representative of the continually responding population over time, as opposed to the full target population.

Examining the estimates of key domains provides insight on whether the potential for bias due to nonresponse error is adversely impacting the survey estimates. To account for nonresponse and ensure the respondent population represents the target population in size, nonresponse weighting adjustments are made to the respondent population. Comparison between the three sets of estimates - (1) base weights for the sample, (2) base weights for the respondents, and (3) nonresponse adjusted weights for respondents - gives an indication of the potential for nonresponse bias and shows how much nonresponse bias is reduced through an appropriate weighting adjustment. This examination will provide insight on whether the NSCG weighting adjustments are appropriately meeting the NSCG survey estimation goals. For example, consider we are attempting to estimate the proportion of the college-educated population that was under 29 years old. Comparisons can be made among the following three sets of estimates:

- Among the 2017 NSCG new sample *cases*, the proportion of college-educated individuals that were under 29 years old was 10.4% (using base weights).
- Among the 2017 NSCG new sample *respondents*, the proportion of college-educated individuals that were under 29 years old was 8.4% (using base weights). These cases had a weighted response rate of 52.8% compared to 65.4% overall which led to an underestimate of 8.4% among respondents compared to the estimate of 10.4% among all sample cases.
- Among the 2017 NSCG new sample *respondents*, the proportion of college-educated individuals that were under 29 years old was 10.4% (using the nonresponse adjusted weights) which provides evidence that nonresponse weight adjustment accounted for the differential nonresponse for this age group.

This pattern is repeated throughout, with lower responding groups being underrepresented and higher responding groups being overrepresented. However, the nonresponse weight adjustment does a good job accounting for this differential response by subgroup and brings the weighted respondent distributions closer to the sample distributions.

R-indicators and corresponding standard errors will be provided for each of the four originating sources of sample for the 2019 NSCG (namely, the 2011 ACS, 2013 ACS, 2015 ACS, and 2017 ACS). R-indicators are useful, in addition to response rate and domain estimates, for assessing the potential for nonresponse bias. R-indicators are based on response propensities calculated using a predetermined balancing model (“balancing propensities”) to provide information on both how different the respondent population is compared to the full sample population, as well as which variables in the predetermined model are driving the variation in nonresponse.

Error Resulting from ACS and NSCG Response Inconsistency

Information from the ACS responses is used to determine NSCG eligibility and to develop the NSCG sampling strata. Inconsistency between ACS responses and NSCG responses has the potential to inflate non-sampling error in multiple ways and will be investigated as part of the 2019 NSCG non-sampling error evaluation. Since we use ACS responses to define the NSCG sampling strata, and we have different sampling rates in each of the strata, inconsistency with NSCG responses on the stratification variables leads to a less efficient sample design with increased variances. For example, we sample non-science and engineering (non-S&E) occupations at much lower rates than S&E occupations which leads to large weights for non-S&E cases and small weights for S&E cases. If a case is identified as non-S&E on the ACS but lists an S&E occupation on the NSCG, then this case with a large weight is introduced into the S&E domain thus increasing the variance of estimates for the S&E domain. The mixing of cases from different sampling strata due to ACS/NSCG response inconsistency thus leads to an inefficient design and contributes to larger variances.

Another opportunity for ACS/NSCG inconsistency leading to non-sampling error is with off-year estimation.⁶ To the extent ACS responses are inconsistent with NSCG responses, using the ACS data to produce estimates for the college-educated population will lead to biased estimates. Therefore, consistency between the ACS and NSCG responses is very important if we want to consider the possibility of producing off-year estimates with smaller bias.

3. METHODS TO MAXIMIZE RESPONSE

To maximize the overall survey response rate, NCSES and the Census Bureau will implement procedures such as conducting extensive locating efforts and collecting the survey data using three different modes (mail, web, and CATI). The contact information obtained for the sample members from the 2017 NSCG and the 2017 ACS will be used to locate the sample members in 2019.

Respondent Locating Techniques

The Census Bureau will refine and use a combination of locating and contact methods based on the past surveys to maximize the survey response rate. The Census Bureau will utilize all available locating tools and resources to make the first contact with the sample person. The Census Bureau

⁶ Off-year estimation would provide estimates for the college-educated population, using only ACS data, in the years where the NSCG is not in the field. For example, as the NSCG is conducted in 2015, 2017, and 2019, off-year estimation would produce estimates for the college-educated population in 2016 and 2018.

will use the U.S. Postal Service's (USPS) automated National Change of Address (NCOA) database to update addresses for the sample. The NCOA incorporates all change of name/address orders submitted to the USPS nationwide and is updated at least biweekly.

The Census Bureau will use the "Return Service Requested" option for mailings to ensure that the postal service will provide a forwarding address for any undeliverable mail. For the majority of the cases, the initial mailing to the NSCG sample members will be a letter introducing the survey and inviting them to complete the survey by the web data collection mode. For the cases that stated a preferred mode for use in future survey rounds (e.g., mailed questionnaire or telephone), NCSES will honor that request by contacting the sample member using the preferred mode to introduce the survey and request their participation.

The locating efforts will include using such sources as educational institutions, alumni associations, and other publicly available data found on the internet, Directory Assistance for published telephone numbers, and FastData and/or Accurint for address and telephone number searches.

Data Collection Methodology

A multimode data collection protocol will be used to improve the likelihood of gaining cooperation from sample cases that are located. Using the findings from the 2010 NSCG mode effects experiment and the positive results of using the web first approach in the 2013, 2015 and 2017 NSCG data collection efforts, the majority of the 2019 NSCG sample cases will initially receive a web invitation letter encouraging response to the survey online. Nonrespondents will be given a paper questionnaire mailing and will be followed up in CATI. The college graduate population is mostly web-literate and, as shown in the 2010 mode effects experiment, the initial offering of a web response option appeals to NSCG potential respondents.

Motivated by the findings from the incentive experiments included in the 2010 and 2013 NSCG data collection efforts, and the positive results from the 2015 and 2017 NSCG incentive usage, NCSES is planning to use monetary incentives to offset potential nonresponse bias in the 2019 NSCG. We plan to offer a \$30 prepaid debit card incentive to a subset of highly influential new sample cases at week 1 of the 2019 NSCG data collection effort. "Highly influential" refers to the cases that had large sampling weights and a low response/locating propensity. We expect to offer \$30 debit card incentives to approximately 13,400 of the 67,000 new sample cases included in the 2019 NSCG. In addition, we will offer a \$30 prepaid debit card incentive to past incentive recipients at week 1 of the 2019 NSCG data collection effort. We expect to offer \$30 debit card incentives to approximately 11,000 of the 81,000 returning sample members. These debit cards will have a six-month usage period at which time the cards will expire and the unused funds will be returned to the Census Bureau and NCSES.

Within the 2019 NSCG data collection effort, the following steps will be taken to maximize response rates and minimize nonresponse:

- Developing "user friendly" survey materials that are simple to understand and use;
- Sending attractive, personalized material, making a reasonable request of the respondent's time, and making it easy for the respondent to comply;
- Using priority mail for targeted mailings to improve the chances of reaching respondents and convincing them that the survey is important;

- Devoting significant time to interviewer training on how to deal with problems related to nonresponse and ensuring that interviewers are appropriately supervised and monitored; and
- Using refusal-conversion strategies that specifically address the reason why a potential respondent has initially refused, and then training conversion specialists in effective counterarguments.

Please see Appendix F for survey mailing materials and Appendix G for the data collection pathway that provides insight on when the different survey mailing materials will be used throughout the data collection effort. Each cycle, the NSCG asks respondents to report in which mode (web, paper, or CATI) they would prefer to complete future rounds of the survey. In past cycles, respondents who reported a preference for paper received a paper questionnaire in the first week of data collection. In the 2017 cycle, about 21% of returning respondents had indicated a preference for paper, but only half of them used the paper questionnaire to respond, while 43% responded online. Given the high rate of web response among people who reported a paper preference, we will be delaying the paper questionnaire until the second week to see if we can maximize the number of web respondents. This delay will still provide respondents with their preferred mode but will attempt to move more people to the web, resulting in faster data processing and less postage paid in questionnaire returns. Additionally, if the web response rate among respondents with a paper preference increases significantly, it may be possible in future NSCG survey cycles to either move the paper questionnaire to week 5, where there will be a large reduction in the number of questionnaires mailed, or remove the mode preference altogether, which would simplify operations in the future.

4. TESTING OF PROCEDURES

Survey Methodological Experiments

Two survey methodological experiments are planned as part of the 2019 NSCG data collection effort. Together, these experiments are designed to help NCSES and the Census Bureau strive toward the following data collection goals:

- Lower overall data collection costs
- Decrease potential for nonresponse bias in the NSCG survey estimates
- Increase or maintain response rates
- Increase efficiency and reduce respondent burden in the data collection methodology

The two methodological experiments are:

- Adaptive Survey Design Experiment
- Mailout Strategy Experiment

An overview of ongoing and past experiments is provided in Supporting Statement A, Section 8; details on the two 2019 experiments can be found in Appendices H and I. Both experiments are planned for both the new sample and the returning sample data collection efforts. This section

introduces the design for each experiment, describes the research questions each experiment is attempting to address, and includes information on the sample selection proposed for these studies.

Adaptive Survey Design Experiment

Beginning in 2013, the NSCG has included adaptive survey design (ASD) experiments in production data collection. Each survey round (2013, 2015, and 2017) has focused on different operational or methodological goals. The 2013 NSCG experiment functioned as a proof of concept. The experiment was relatively small (4,000-case treatment and control groups) and only used new sample members. From an operational perspective, the experiment was a success, as the NSCG team was able to work with several Census data collection systems to enable adaptive interventions. A team of methodologists from Census and NCSES would determine which cases should be subject to interventions, that information was provided to programming staff, and *ad hoc* intervention files were manually created and transmitted to the various systems. The 2015 and 2017 adaptive design experiments also relied upon the same intervention file creation and transmission mechanisms.

The 2013 experiment used variable-level and category-level unconditional partial R-indicators to identify which subgroups should be included in enhanced or reduced contact effort interventions. Every two weeks, survey managers and methodologists would meet to finalize the cases selected for intervention. Overall R-indicators, as well as overall and subgroup-level response rates, were monitored to understand the effect of targeted interventions on high-level progress and quality metrics. The overall size of the 2013 experiment groups meant that intervention groups were very small (often less than 100 cases), so we did not find statistically significant differences. However, promising trends in both response rate and R-indicators (along with a 15% lower cost per case in the treatment vs the control) led to an expanded experiment in 2015.

The 2015 experiment was similar to the 2013 experiment, except it included both new and returning sample members, and the experimental groups were much larger than in 2013. Among new sample members, the treatment and control groups each had 8,300 cases, and among returning sample members the treatment and control groups each had 10,000 cases. The main goal in 2015 was to apply fewer, larger interventions in a way that would improve representativeness in the treatment group versus the control group, while monitoring the effect on response rate and data collection cost. Interventions made during the 2015 experiment resulted in statistically significant improvements in overall representativeness and subgroup level variability in the treatment groups of both new and returning sample members, when compared to the control groups. Additionally, the returning sample members had a significantly higher response rate, while the new sample members showed no significant reduction in response rate. Overall costs per case were slightly lower in the treatment groups than in the control groups. Counterfactuals also illustrated the fact that the treatment groups achieved the final representativeness level of the control groups after spending only 60% of the control groups' data collection costs. These findings have been submitted to the *Journal of Survey Statistics and Methodology* (Coffey, et al, 2018)⁷ for publication as they are strong support for the potential of adaptive survey design to be able to improve cost and quality outcomes in the current survey climate.

Given the methodological success of the 2015 experiment, the 2017 adaptive design experiment was designed to increase automation of the case identification and file preparation process. Prior to 2017, nearly all aspects of the adaptive design experiment, from ingesting and manipulating paradata and metric creation to creation and delivery of intervention files, required human effort and

⁷ Coffey, S., Reist, B., Miller, P. (2018). Interventions on Call: Dynamic Adaptive Design in the National Survey of College Graduates. Under Review in *Journal of Survey Statistics and Methodology*.

interaction. Very few processes were automated on a schedule. While this provided flexibility if additional analytic questions of interest arose, it required a disparate group of individuals to all be available on “intervention days” in order to submit intervention file requests, create intervention files, translate those into the correct input files, and deliver those files.

In addition to the increase in automation in 2017, the number of data quality metrics that were monitored during data collection also increased to help understand the relationship between them and how they reacted to data collection interventions. Along with response rates and R-indicators, which have been utilized since the 2013 experiment, we also calculated and monitored the coefficients of variation (CVs) of the balancing propensities and response propensities. Additionally, we monitored the completion rate for the “high effort” cases in the treatment and control groups to see if the interventions were increasing response among under-represented cases.

Adaptive Survey Design in the 2019 NSCG

In the 2019 NSCG, NCSES and the Census Bureau are continuing to innovate with respect to adaptive survey design. In past rounds, we used variable-level partial unconditional R-indicators in order to identify cases that fell into very under-represented or very over-represented subgroups. These classifications were defined by values of the R-indicator itself. When compared to the production data collection pathway (the control), under-represented cases received more contact effort, and over-represented cases received less contact effort. However, when making these interventions, we did not attempt to predict the effect on survey outcomes, such as the expected gain in response, reduction in mean squared error of survey estimates, or effects on survey costs by applying an intervention on a particular subset of cases.

In other words, the cost savings and relatively stable response rates our experiments have achieved was not by design – we did not control for those outcomes, though we did monitor them. However, it is possible that we could have obtained better results by ensuring that our interventions maintained a neutral cost per case while intervening on a larger group of cases (or reducing contact effort on fewer cases). This may have resulted in further improved survey outcomes. There is also no guarantee that, in the future, continuing to intervene in the same way will result in the same outcomes of cost savings and stable response rates. Further, while representativeness is often considered a proxy for nonresponse bias, the interventions made historically do not take into account any expected effect on key estimates, such as the mean squared error or the size of nonresponse adjustments.

To address these research gaps, we propose an adaptive survey design experiment that focuses on improving a data quality metric (e.g., root mean squared error, or RMSE) for a given cost to see if we can also apply interventions in a way that improves the quality of survey estimates directly, rather than improving proxies, like R-indicators. The work will include prediction of response propensity, expected costs, and both item- and unit-level imputation under different data collection scenarios in order to predict key outcomes of interest. High quality predictions will allow us to selectively apply data collection features in a way that has a direct impact on key survey estimates in the NSCG. To evaluate the effect of adaptive design on the key survey estimates of the NSCG, the following high-level methodology will be used:

- (1) After each week of data collection, obtain posterior estimates of response propensity for all unresolved cases, using the methodology in Wagner and Hubbard (2014).⁸

⁸ Wagner, J., Hubbard, F. (2014). Producing Unbiased Estimates of Propensity Models During Data Collection, *Journal of Survey Statistics and Methodology*, 2:3, Pages 323–342, <https://doi.org/10.1093/jssam/smu009>

(2) For pre-identified key survey variables, use available paradata and sampling frame information to impute missing values for all unresolved cases. The resulting estimate of the mean for each of these variables will be treated as the unbiased target parameters of interest.

(3) For given response propensity cut-points identified using the propensity output in (1), assign cases to adaptive interventions (e.g., sending cases to CATI early, or replacing questionnaires with web invites) and re-compute the estimates of response propensity from step (1) under the new scenario to predict who will respond.

(4) Re-compute estimates from step (2) after removing cases who are predicted to be nonrespondents from the dataset (meaning fewer unresolved cases will be imputed). This will enable a comparison of the estimates in step (3) vs step (2). The root mean squared error (RMSE) will be estimated using the output of step (2) as the truth.

(5) Compare the estimated cost of contacting all unresolved cases using the current data collection protocol versus only those retained in steps (3) and (4) using a new data collection protocol (e.g., introducing CATI early, withholding different contact strategies, etc.). Cost estimates will be obtained from prior rounds of the NSCG, incorporating information from the sampling frame, and costs by mode of response.

(6) Repeat steps (3) - (5) throughout data collection, plotting the estimated costs versus the RMSE by cut-point. Times where there are sharp drops in cost with small increases in RMSE point to times where design features should be changed to reduce cost without increasing the expected RMSE.

To leverage historical information that can aid in predicting response propensity, along with survey outcomes and cost, the predictive models developed here will use a Bayesian framework. Historical predictors are used to create *priors*, or initial predictive model coefficients, which are then updated with the current round of information, as more data is collected to create *posteriors*. The posteriors are then used in the predictive models mentioned above. A benefit of these types of models, particularly early in data collection, is that the historical data can compensate for the data that has not been collected in the current survey round.

The Census Bureau is currently building and evaluating these predictive models, as well as identifying the survey conditions we would want to simulate (e.g., neutral cost-per-case, 10% reduction, 10% increase) from historical data. This will help both Census and NCSES understand how these decisions would be made during data collection. In the event that these models do not have sufficient power, or result in poor predictions of survey outcomes, we plan to adapt steps (1) to (6) above in order to maximize the survey R-indicator for a given cost constraint. The propensity models underlying the NSCG R-indicators have been validated over the past three data collection cycles and include variables that are highly correlated with survey outcomes,⁹ giving us confidence that our interventions are improving data collection, even if the metric we use is only a proxy for nonresponse bias. The Census Bureau, in conjunction with NCSES, plans to make final decisions about the models used for intervention by December 2018. This ensures the intervention methods selected have been jointly reviewed and agreed upon before data collection begins.

The Census Bureau is not alone in exploring the use of Bayesian models in adaptive survey design. The Census Bureau is working in consultation with Dr. Michael Elliott, from the University of

⁹ Singer, P., Sirkis, R. (2013). "Evaluating the Consistency between Responses to the 2010 NSCG and the 2009 ACS," Proceedings of the Survey Research Methods Section, American Statistical Association. 2013.

Michigan, who has expertise with both Census data and Bayesian models (Elliott, et al., 2000¹⁰; Elliott, 2017¹¹). Additionally, ongoing research is being conducted by survey methodologists and statisticians from the Institute for Social Research at the University of Michigan as well as Statistics Netherlands. A paper on the topic by researchers from both institutions was recently published (Schouten, et al, 2018).¹²

In preparation for the 2015 NSCG adaptive design experiment, and in response to a request from OMB, NCSES and the Census Bureau prepared a detailed adaptive design plan that included the time points where potential interventions would occur and the interventions that were available at each time point. This documentation served us throughout the 2015 and 2017 NSCG data collections and, given the similarities between those implementations and the 2019 NSCG production data collection pathway, this documentation will continue to serve as an overview of the adaptive design approach for the 2019 NSCG, with appropriate adjustments incorporated to reflect changes in the production data collection operations plan.

Appendix H includes a table outlining the production data collection mailout schedule and the potential interventions available at each time point. Additionally, Appendix H discusses the adaptive design goals, interventions, and monitoring metrics considered for the experiment.

Mailout Strategy Experiment

In the 2017 NSCG cycle, NCSES experimented with a new set of mailing materials. This experiment showed that by updating the type and content of mailing materials, the number of mail contacts could be reduced without harming response rates, sample representativeness, or key estimates. Specifically, the majority of the NSCG’s mailing materials use standard letter-sized white envelopes, with the exception of a letter-sized brown envelope at the end of data collection and the questionnaire packets. Focus group research suggested that respondents are more likely to discard future letters if they look the same as one they have already opened because they know what it contains.¹³ Therefore, the 2017 experiment used a variety of envelopes, including an oversized white envelope similar to what is used in the American Community Survey, a pressure-sealed letter-sized envelope, and a tabbed postcard. Additionally, for the returning sample cases, emails were sent immediately following mail contacts instead of as stand-alone reminders because research shows that simultaneous reminders in multiple modes are more effective than sequential reminders.¹⁴

While the new mail materials were successful in generating response with fewer contacts, the experiment tested one overall strategy. It did not have the sample size available to test variations of that strategy. The 2019 NSCG mailout strategy experiment is designed to test variations of the strategy implemented in the previous survey cycle with a specific focus on the use of due dates in survey contact materials.

¹⁰ Elliott, M.R., Little, R.J.A, and Lewitzky, S. (2000). Subsampling Callbacks to Improve Survey Efficiency. *Journal of the American Statistical Association*, 95, 730-738.

¹¹ Elliott, M.R. (2017). Preliminary ideas for responsive design implementation. Internal Census Bureau – Center for Adaptive Design Memo. (2/27/2017).

¹² Schouten, B., Mushkudiani, N., Shlomo, N., Durrant, G., Lundquist, P., Wager, J. (2018). A Bayesian Analysis of Design Parameters in Survey Data Collection. *Journal of Survey Statistics and Methodology*, <https://doi.org/10.1093/jssam/smy012>.

¹³ Morales, G. D. (2016). *Assessing the National Survey of College Graduates Mailing Materials: Focus Group Sessions*.

¹⁴ Dillman, D., Smyth, J., & Christian, L. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method (4th Edition)*. New York: Wiley & Sons.

Feedback from focus groups suggested that providing a specific due date for the survey makes the request seem more urgent as compared to “please respond within two weeks,” which is currently used in the NSCG.¹⁵ The 2017 experiment used a due date two weeks from the initial invitation (May 4th) but, given the 6-month data collection period, it is not clear that this was the most effective way to present the due date. To answer these questions, the 2019 mailout strategy experiment has two treatments:

1. location of the due date; and
2. how and when the date is presented.

The examination of these two different strategies creates a total of four treatment groups that will be included within this experiment as displayed in the following table. The experiment will be conducted separately for the new sample cases and the returning sample cases.

Deadline Location	Deadline Timing	Treatment
Within Letter	Initial Mailout (Weeks 1 and 2)	LI (Letter/Initial)*
	Late Mailout (Week 12)	LL (Letter/Late)
Envelope and Letter	Initial Mailout (Weeks 1 and 2)	EI (Envelope/Initial)
	Late Mailout (Week 12)	EL (Envelope/Late)

*Control Group

Appendix I provides more detail about the different treatments, the rationale for their inclusion in the mailout strategy experiment, and the research questions we are attempting to answer through this experiment.

Designing the Sample Selection for the 2019 NSCG Methodological Experiments

Two methodology studies are proposed for the 2019 NSCG: the adaptive design experiment and the mailout strategy experiment. This section describes the sample selection methodology that will be used to create representative samples for each treatment group within the two experiments. The eligibility criteria for selection into each of the studies are as follows:

- Adaptive Survey Design Experiment: All cases are eligible for selection
- Mailout Strategy Experiment: Cases must not be selected in the adaptive design treatment or control group, must have a mailable address, and must not have been sent new mailing materials as part of the 2017 NSCG contact strategies experiment (this last criterion applies to 2019 old cohort cases only).

The sample for the adaptive design experiment will be selected independently of the sample selection for the mailout strategy experiment. Keeping the adaptive design cases separate from the other experiment will allow maximum flexibility in data collection interventions for these cases. In addition, the sample selection will occur separately for the new sample cases and the returning sample cases. This separation will allow for separate analysis for these two different sets of potential respondents. The main steps associated with the sample selection for the 2019 NSCG methodological studies are described below.

Step 1: Identification and Use of Sort Variables

¹⁵ Morales, G. D. (2016). *Assessing the National Survey of College Graduates Mailing Materials: Focus Group Sessions*.

Since the samples for the treatment and control groups within the methodological studies will be selected using systematic random sampling, the identification of sort variables and the use of an appropriate sort order is extremely important. Including a particular variable in the sort ensures similar distributions of the levels of that variable across the control and treatment groups.

Incentives are proposed for use in the 2019 NSCG. It has been shown in methodological studies from previous NSCG surveys that incentives are highly influential on response. An incentive indicator variable will be used as the first sort variable for both methodological studies. The 2019 NSCG sample design variables are also highly predictive of response and will also be used as sort variables in all studies. The specific sort variables used for each experiment are:

- Incentive indicator
- 2019 NSCG sampling cell and sort variables

Step 2: Select the Samples

For the new sample adaptive design experiment, a systematic random sample of approximately 8,000 cases will be selected to the treatment group and 8,000 cases will be selected to the control group. For the new sample mailout strategy experiment, the sample will be subset to the eligible population (see above for eligibility criteria) and a systematic random sample of approximately 6,875 cases will be selected in each of three treatment groups. All eligible new sample cases not selected into the adaptive design treatment, adaptive design control, or mailout strategy treatment groups will be assigned to the mailout strategy control group (approximately 28,000 cases).

For the returning sample adaptive design experiment, a systematic random sample of approximately 10,000 cases will be selected to the treatment group and 10,000 cases will be selected to the control group. For the returning sample mailout strategy experiment, the sample will be subset to the eligible population (see above for eligibility criteria) and a systematic random sample of approximately 6,292 cases will be selected into each of three treatment groups. All eligible returning sample cases not selected into the adaptive design treatment, adaptive design control, or mailout strategy treatment groups will be assigned to the mailout strategy control group (approximately 30,000 cases).

Minimum Detectable Differences for the 2019 NSCG Methodological Experiments

Appendix J provides information on the minimum detectable differences achieved by the sample sizes associated with the 2019 NSCG methodological experiments.

Analysis of Methodological Experiments

In addition to the analysis discussed in the sections describing the experiments, we will calculate several metrics to evaluate the effects of the methodological interventions and will compare the metrics between the control group and treatment groups. We will evaluate:

- response rates (overall and by subgroup);
- R-indicators (overall R-indicators, variable-level partial R-indicators, and category-level partial R-indicators);
- mean square error (MSE) effect on key estimates; and
- cost per sample case/cost per complete interview (overall and by subgroup).

The subgroups that will be broken out are the ones that primarily drive differences in response rates and include: age group, race/ethnicity, highest degree, and hard-to-enumerate.

5. CONTACTS FOR STATISTICAL ASPECTS OF DATA COLLECTION

Chief consultant on statistical aspects of data collection at the Census Bureau is Stephen Simoncini, NSCG Survey Director (301) 763-4816. The Demographic Statistical Methods Division will manage all sample selection operations at the Census Bureau.

At NCSES, the contacts for statistical aspects of data collection are Samson Adeshiyan, NCSES Chief Statistician (703) 292-7769, and Lynn Milan, NSCG Project Officer (703) 292-2275.