PART B. DESCRIPTION OF STATISTICAL METHODOLOGY

B.1.   Statistical Design and Estimation

B.1.1   Survey Population

The Facilities survey is designed to provide national estimates for U.S. colleges and universities with science and engineering research expenditures equal to or greater than $1 million in the prior academic fiscal year (i.e., in FY 2018 for the FY 2019 cycle and FY 2020 for the FY 2021 cycle). The FY 2019 cycle is anticipated to be a census of approximately 600 institutions. The listing of eligible institutions will be derived from the NCSES Survey of Higher Education Research and Development (HERD). No sampling will be conducted. The response rate on the FY 2017 survey was 98%.

B.1.2   Estimation Procedures

No sampling weights will be required because the survey is a census. However, adjustments will be performed for both unit nonresponse and item nonresponse, with the approach depending on the level of nonresponse and the characteristics of the particular item involved (for item nonresponse). For FY 2017, combining both unit and item nonresponses, the missing response rates were very low (less than 4% for all items) and were primarily attributed to unit nonresponse.

Adjustments for Unit Nonresponse

Since some nonresponse is likely, provisions will be made to compensate for the missing data in the survey estimates. Unit nonresponse (an institution does not respond to the entire survey) occurs when there is no information for a sampled unit, most often because of refusal to participate in the survey.

In the FY 2017 survey cycle, unit nonresponse for the survey was handled by imputing for the missing items in the unit. This procedure will be followed for the FY 2019 and FY 2021 survey cycles. Procedures for item nonresponse are detailed below.

Adjustments for Item Nonresponse

Item nonresponse occurs when there is no information for a respondent on an individual item in the questionnaire, most often because of refusal to answer that item or the provision of an invalid response (e.g., one that falls outside of the possible range of values). We will use imputation on selected variables to adjust for item nonresponse.

The imputation approach uses multivariate regression models, including linear and/or logistic regression models. A special feature of the survey data that is addressed by the imputation methods is the presence of legitimate zero values that frequently occur for some items (e.g., questions with S&E field categories). For example, institutions that report repair and renovation costs in some fields (Question 8), usually report zero costs in other fields. Analysts may wish to examine data such as the number of institutions that have repair and renovation projects in a specific field (e.g., agricultural sciences), as well as the total cost of these projects. This type of analysis is supported if the imputed data follow the same pattern of zero and nonzero responses as the reported data. To maintain this pattern of item response in the imputation process, a logistic regression is applied to decide whether the imputed value should be zero or not. For the cases to be imputed with nonzero (positive) values, a linear regression is then conducted to impute the exact value.

Generally, the imputation is done in the order in which each item appears in the questionnaire except in a few cases. One example of not following the order of the questionnaire is for the completion cost and the total net assignable square feet (NASF) for the portion of a new construction project used for S&E research, by field (Question 10E). This item is completed prior to the imputation of Question 10C and Question 10D, which ask about the entire project's gross square feet and completion cost. The imputation of both Question 10C and Question 10D are based on the imputed Question 10E.

For all survey items, the predictor variables are examined to make sure that any variable that is needed to preserve proper routing through the questionnaire or consistency with other survey data is taken into account. Due to stability in the questionnaire, the same predictor variables have been used for imputation in the last eight survey cycles. For all imputed items, a set of standard predictor variables (i.e., core predictors) is used in the regression models for imputation. The core predictors are:

- Control (public/private);

- Highest degree granted (doctorate/nondoctorate);

- Existence of a medical school (yes/no);

- Total R&D expenditures in S&E from the previous year (i.e. from FY 2018 for the FY 2019 survey); and

- Total NASF of research space (total across S&E fields).

Other than total NASF (computed from Question 2) and existence of a medical school (based on data from the American Association of Medical Colleges and the American Association of Colleges of Osteopathic Medical Schools), these predictors are obtained from the NCSES HERD data file.

For some items, other correlated variables are also included in the models as additional predictor variables, in a manner consistent with the imputation from previous cycles. For the FY 2017 survey, responses from the preceding survey cycle (FY 2015) were used to generate a second model for the imputation of Questions 2, 3, 4, 5, and 8. This additional model included the FY 2015 Facilities survey responses to the item as an additional predictor, and was therefore generated using only the institutions that responded for the item in both cycles. For institutions that did not respond to an item during the FY 2015 cycle, the standard model (i.e. without the FY 2015 data) was used. The second model generally resulted in improved imputation for the items where it was used. For example, using the R^2 to compare the goodness of fit for the two imputation models for Question 3, the standard model R^2 was 0.743, which increased to 0.925 after adding the FY 2015 data to the second model. This procedure of generating the additional model using previous cycle data, where available, will be followed for the FY 2019 and FY 2021 survey cycles.

In the imputation models for some items, rather than imputing the item itself, a ratio of the item with another item is imputed. This is done because the ratio is much more stable and predictable than the item itself. In this situation, the ratio is the outcome variable in the linear regression model. The imputed ratio is then used to calculate the imputed item value.

For all imputed items an influence statistic (DFFITS) is calculated for each institution to identify outliers in the regression model. The DFFITS statistic produced in SAS (see SAS 9.4 manual) is a scaled measure of the change in the predicted value for the ith observation calculated by deleting the ith observation. If a large change of predicted values is caused by deleting a few outliers, then these unusual cases have a large influence on the fitted regression line and the regression prediction (the imputed value) can be distorted. Some outlying cases with extreme values of DFFITS are excluded from the regression models, to minimize distortion.

The predicted values from the regression models are copied into the data file as the imputed responses. The imputation flag indicates when the value was imputed. When imputation is completed, all edit checks that had been done before imputation are run again to identify any data inconsistencies caused by the imputation.

B.2.   Survey Procedures

The Facilities survey is a web survey, with telephone and email follow-up.

The president of each institution is mailed a cover letter, which includes information on accessing a copy of the survey questionnaire on the survey website, and a copy of the Infobrief from the previous (FY 2017) survey cycle. Depending on whether the institution plans to retain the previous survey cycle coordinator for the upcoming survey cycle, the president may also receive one of two institutional coordinator forms (see below). The coordinator acts as the central communication point for NCSES and the contractor collecting the data.

During the previous survey cycles, presidents were asked if they wished to keep that year's institutional coordinator for future data collections. If a president indicated that the coordinator would be the same, the president's cover letter identifies the previous cycle coordinator and indicates that the survey materials will be sent directly to this coordinator. Two days later, the prior cycle's coordinator is sent a letter by email indicating that data collection is beginning and that his or her name has been provided to the president as the past coordinator. At this time, the coordinator also receives a copy of all survey materials.

If the institution was in the previous survey cycle but the president did not indicate at that time whether he/she wished to keep the same coordinator for the next cycle's data collection, the president receives a pre-filled form identifying the previous cycle coordinator. This coordinator identification form asks the president to indicate whether he/she wishes to continue with the same coordinator or name a new coordinator.

If the institution was not in the previous survey cycle, he/she receives a blank coordinator identification form to use to indicate the current cycle's coordinator.

For presidents' offices that receive a coordinator identification form, if no response is received by the coordinator identification due date indicated in the letter, telephone prompts are used to determine the name and contact information for an institutional coordinator. Following designation of the coordinator, the coordinator is notified that he or she has been appointed survey coordinator and provided with the survey materials by email. See Attachment D for draft contact materials.

Regular email and/or telephone prompts are used to encourage the institution to respond. Institutions have the option of completing either a paper copy of the questionnaire or providing the data on the web through a designated web site. Based on past experience, we expect 99% of the responding institutions to report using the web. Returned questionnaires are examined for quality and completeness using computerized edits and visual inspections by the contracting staff. In the case of questionnaires completed on the web, computerized edits check for quality and completeness as the data are entered, and prompt the respondents if problems are found. If key items have missing data or other problems appear in the data (e.g., two responses appear to be inconsistent), then respondents are contacted again to resolve the issues.

B.3.   Methods for Maximizing Response Rates

A key to achieving a response rate in line with recent surveys is tracking the response status of each institution, with telephone, email, and mail follow-up of those institutions that do not respond in a timely manner. The survey responses will be monitored through an automated receipt control system. Approximately one week after the initial survey invitation email is sent, coordinators who have not replied to verify receipt will be called by the contractor to verify receipt of the email and gain cooperation. Additional telephone, email, or mail prompts will be made as the data collection period continues.

Several other steps will be taken to maximize the response rate. The survey materials will provide a toll-free number that people may call to resolve questions about the survey. Respondents may seek help by email. In addition, standard survey techniques that have proven successful in other academic survey efforts will be employed to achieve a maximum response rate. These techniques include:

- A cover letter signed by the NCSES director.

- Institutional coordinators will be contacted by telephone prior to the conclusion of the survey. This contact is intended both to offer assistance to respondents and to encourage their speedy response.

- Follow-up telephone calls and emails will be made, and letters sent to nonresponding institutions as required. These follow-up contacts have been demonstrated to achieve significant improvements in response rates.

Finally, institutions will be informed in their materials that institution-level survey responses are currently available for the previous survey cycles and institutional responses will also be available for the current FY 2019 (and FY 2021) survey. These data will be available on a publicly accessible database on the NCSES website. NCSES believes that having publicly available data will maximize responses rates because institutions will be more likely to participate if they believe the data will be useful to them.

B.4.   Tests of Procedures and Methods

The questionnaire is based on versions of the survey used in previous cycles. As part of survey improvement efforts, the survey staff participates in an extensive debriefing after each survey cycle. During the debriefing the staff discusses issues such as the questions respondents ask most frequently, the survey questions that posed problems for respondents, any administrative issues that arose, and other survey improvement issues. In addition, the survey paradata are analyzed after each cycle implementation. The paradata include: a list of the "other specific responses" for each question, the frequency of error messages for each question, missing data, consistency of question responses, interviewer logs of conversations with respondents, comments reported by respondents on their survey responses, and completed survey return flow. Based on these analyses, survey questions or procedures may be revised.

B.5.   Individuals Responsible for Study Design and Performance

The individuals listed below participated in the study design.

Jock Black, NCSES	703-292-7802
Michael Gibbons, NCSES	703-292-4590
John Jankowski, NCSES	703-292-7781
Rebecca Morrison, NCSES	703-292-7794
Lucinda Gray, Westat	301-251-4336

Eric Jodts, Westat   301-610-8844

Feven Negga, Westat   240-314-2335

Mina Muller, Westat   301-738-8319

The contractor for the FY 2019 and FY 2021 data collection is Westat. Michael Gibbons at NCSES is the contracting officer's representative for the contract.

Attachments:

    A. NSF Act of 1950 and America COMPETES Reauthorization Act of 2010
    B. FY 2019 Facilities Survey questionnaire
    C. Federal Register Notice 84 FR 9840
    D. Draft contact materials for FY 2019 Facilities Survey

Attachment A

NSF Act of 1950
and
America COMPETES Reauthorization
Act of 2010

Attachment B

FY 2019 Facilities Survey Questionnaire

Attachment C

Federal Register Notice 84 FR 9840

Attachment D

Draft Contact Materials for FY 2019 Facilities Survey