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 research expenditures equal to or greater than $1 million in the prior academic fiscal year (i.e., in FY 2014 for the FY 2015 cycle and FY 2016 for the FY 2017 cycle). The FY 2015 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. No sampling will be conducted. The response rate on the FY 2013 survey was 99%.

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).

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 2013 survey cycle, unit nonresponse for the research space section of the survey was handled by imputing for the missing items in the unit. This procedure will be followed for the FY 2015 and FY 2017 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 7), 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 agricultural science, 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 9E). This item is completed prior to the imputation of Question 9C and Question 9D, which ask about the entire project's gross square feet and completion cost. The imputation of both Question 9C and Question 9D are based on the imputed Question 9E.

For items with low rates of missing data (defined as less than 5% missing data in FY 2013), a set of standard predictor variables (i.e., core predictors) is used in the regression models for imputation. Additional predictor variables are used when the survey items have higher rates of nonresponse. These predictors are chosen by examining the data to find correlated variables. For FY 2013, the survey items with high missing rates (5% or greater) were items on the source of project funding for new construction projects, Question 10A2, Question 10B2, Question 10C2 and Question 10 Total 2. Except for Question 10 Total 2 (which is filled in with total funds for new construction projects in Question 9E), logistic and/or linear regression models were applied to these items. In addition to the

core predictors, the model predictors include all the variables related to funding for new construction, including field-specific variables from Question 9E. For all survey items, the predictor variables are examined to make sure that any variable that was 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 five survey cycles. For FY 2013, combining both unit and item nonresponses, the missing response rates were still low (less than 5% for all except four items). The core predictors were:

- Control (public/private);
- Highest degree granted (doctorate/nondoctorate);
- Existence of a medical school (yes/no);
- FY 2012 total R&D expenditures in S&E; 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 NSF HERD data file.

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 items with higher missing rates, an influence statistic (DFFITS) is calculated for each institution to identify outliers in the regression model. The DFFITS statistic produced in SAS (see SAS 9.2 manual) is a scaled measure of the change in the predicted value for the *i*th observation calculated by deleting the *i*th 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 also excluded if the missing rate is less than 2%.

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 mixed-mode mail and web survey, with telephone and email follow-up.

The president of each institution is mailed a copy of the questionnaire, a cover letter, and a copy of the report from the previous (FY 2013) survey cycle. In addition, the president receives one of two institutional coordinator forms depending on whether the institution planned to retain the previous survey cycle coordinator for the upcoming survey cycle (see below). The coordinator acts as the central communication point for NCSES and the contractor collecting the data.

During the previous survey cycle each president is asked if they wish to keep that year's institutional coordinator for the next cycle's data collection. If the president indicated that the coordinator would be the same, the president is sent a pre-filled form along with the previously mentioned materials indicating their current cycle coordinator (and providing an opportunity to name a new coordinator if he/she wishes to do so). Simultaneously, the prior cycle's coordinator receives a letter 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 materials.

If the president indicated in the previous cycle's data collection that he/she did not wish to keep the same coordinator for the next cycle's data collection, the president receives a blank form to use to indicate the current cycle's coordinator. To aid a president in selecting a new coordinator, the letter to the president indicates who acted as the coordinator in the previous survey cycle (if the institution responded to the cycle). Simultaneously, the prior cycle's coordinator receives a letter indicating that data collection is beginning, that a letter has been sent to the institution's president requesting a coordinator, 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 materials.

If no response is received from the president's office within a week, 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. See Attachment E 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 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 three weeks after the initial mailing, the contractor will begin calling nonrespondents to verify they received the questionnaire and to prompt response. Additional telephone or email 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 800 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 will be made to nonresponding institutions as required. These follow-up calls are expected 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 2015 (and FY 2017) survey. These data will be available on a publicly accessible database on the NSF 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, 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	240-314-2335
Eric Jodts, Westat	301-610-8844
Feven Negga, Westat	301-251-4336

The contractor for the FY 2015 and FY 2017 data collection is Westat. Michael Gibbons at NSF/NCSES is the contracting officer's representative for the contract.

Attachments:

A. NSF Act of 1950 and America COMPETES Reauthorization Act of 2010
B. Federal Register Notice 79 FR 35577
C. FY 2015 Facilities Survey questionnaire
D. Federal Register Notice 80 FR 16030
E. Draft contact materials for FY 2015 Facilities Survey