Department of Commerce
U.S. Census Bureau
OMB Information Collection Request
Business Research and Development Survey
(Form BRD-1)
OMB Control Number: 0607-0912

Supporting Statement Part B.   Collection of Information Employing Statistical Methods

1.   **Description of the Universe and Respondent Selection**

The sample frame consists of a list of about 1 million company records that are created by aggregating data from establishment records contained on the Census Bureau's Business Register. The companies represented on the sample frame are located in the United States, are large enough to conduct research and development, are primarily composed of for-profit entities, are within the scope of the survey based on the North American Industry Classification System (NAICS), and are in business at the end of the year prior to the survey year. Each company record on the sample frame is assigned a 6-digit NAICS code, regardless of the number of business activities the company conducts.

A probability sample of approximately 45,000 companies will be selected annually during the clearance period to represent the approximately 1 million companies on the sample frame. For the five years from 2012-2016, the survey's overall response rate averaged 76.6% and we expect it to remain fairly consistent for the 2018-2020 Surveys.

2.   **Procedures for Collection of Information**

Statistical Methodology for Stratification and Sample Selection

The sample design takes advantage of several pieces of information concerning the conduct of R&D. They are (1) reported and imputed R&D information in the Survey over the past 5 years, (2) R&D information from the Company Organization Survey (COS) and the Internal Revenue Service that is extracted from the Census Bureau's Business Register, (3) R&D information from online financial databases, and (4) R&D information from various trade associations. It is estimated that about 3 percent of all U.S. companies have R&D expenses, so additional information about R&D reporting is used to make the sample design more efficient.

The sample frame of approximately 1 million companies is first partitioned into three groups. Group 1 consists of companies where the status of R&D is known to be positive from the pieces of R&D information given above. Group 2 consists of companies where the status of R&D is known to be zero from the pieces of R&D information given above. Group 3 consists of companies where the status of R&D is unknown. Table 1 provides approximate sample frame

counts by group based on information from the 2017 Survey.

Table 1 – Approximate Sample Frame Counts by Group

Group/Category	Number of Company Records
Group 1 – Known Positive R&D status	32,000
Group 2 – Known Zero R&D status	28,000
Group 3 – Unknown R&D status	1,038,000
TOTAL	1,098,000

After the three groups are formed, the sample frame is stratified and companies to be selected with certainty are identified. Groups 1 and 3 are stratified by industry category that reflects the detail expected for the published estimates, but Group 2 is its own stratum and is not further stratified by industry category. An independent probability sample is then selected within each group and industry stratum, using constraints on the coefficient of variation achieved on estimated totals to allocate the target sample size of about 45,000 companies to the strata. In addition, we incorporate constraints on the minimum probability of selection to limit the size of the sample weights.

Estimation Procedure

For a majority of the estimates, an adjusted Horvitz-Thompson (H-T) estimator and variance is computed. For the adjusted H-T estimator, an individual company's data is weighted, and the weights are based on the product of two component weights. The first component weight is the sample weight, which is the reciprocal of the given company's probability of being selected in the sample. The second component weight is the unit nonresponse adjustment factor, which is applied to reported or imputed data for the given company. Then, the weighted company data are aggregated to produce estimated totals.

3.   **Methods to Maximize Response and Account for Nonresponse**

Use of Multiple Modes – Sampled companies have the option to respond either via the web or mail.

Mandatory Reporting Requirement – Title 13, USC, requires businesses and other organizations that receive the questionnaire to answer the questions and return the report to the Census Bureau, which protects the confidentiality of the survey responses.

Follow-up procedures – The due date for companies to report will be 6 weeks from mail out. A due date reminder letter will be sent two weeks prior to the due date. Follow-up letters will be sent in May, June, and August. Depending on response, there will be an optional fourth follow-up package sent in October which would include a duplicate form. In addition, Census Bureau staff will telephone companies among the largest R&D performers that have not returned a survey form or requested a filing time extension by the due date.

These companies account for approximately 85 percent of the value of the data and their responses are critical for the completeness of the estimates.

Estimating for missing data – Unit nonresponse is handled by adjusting weighted reported and imputed data. Item nonresponse for a given company is handled by imputation, which is primarily based on three different methods – (1) direct substitution of data from the company's annual Form 10-K report, if the company is publicly traded, or administrative data for the company; (2) ratio imputation using the company's survey data for both current and prior year; and (3) ratio imputation using current-year survey data from both the company and other companies, which reported both the survey item being imputed for the company, as well as the other survey item used in the ratio.

4.   **Tests of Procedures or Methods**

The NSF and the Census Bureau annually conduct debriefings with survey respondents to assess the effectiveness of current question wording, instructions, and tools. In the event that these debriefings or other respondent feedback suggest the need for substantive changes to survey questions, any such changes will be cognitively tested with respondents prior to implementation.

5.   **Contacts for Statistical Aspects and Data Collection**

Persons responsible for sample design and selection:

Amy M. Newman-Smith, Methodology Director
Manufacturing, Investment, and Construction Programs
Economic Statistical Methods Division
U.S. Census Bureau
(301) 763-6595
(301) 763-4718 (FAX)
amy.m.newman.smith@census.gov

Colt S. Viehdorfer, Mathematical Statistician
Economic Statistical Methods Division
U.S. Census Bureau
(301) 763-6796
(301) 763-4718 (FAX)
david.l.kinyon@census.gov

Jock Black, Mathematical Statistician
National Center for Science and Engineering Statistics
National Science Foundation
(703) 292-7802
(703) 292-9092 (FAX)
jblack@nsf.gov

Persons responsible for data collection:

Richard S. Hough, Assistant Division Chief
R&D and Special Surveys
Economic Reimbursable Surveys Division
U.S. Census Bureau
(301) 763-4823
(301) 763-4718 (FAX)
richard.s.hough@census.gov

Michael Flaherty, Chief
Research, Development, and Innovation Surveys Branch
Economic Reimbursable Surveys Division
U.S. Census Bureau
(301) 763-7699
(301) 763-4718 (FAX)
michael.j.flaherty@census.gov

Persons responsible for analysis of the statistics and publication:

Raymond M. Wolfe, Senior Analyst
Research and Development Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation
(703) 292-7789
(703) 292-9091 (FAX)
rwolfe@nsf.gov

John E. Jankowski, Program Director
Research and Development Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation
(703) 292-7781
(703) 292-9091 (FAX)
jjankows@nsf.gov

Attachments

A. BRD-1 Questionnaire

B. Letters

    1) Initial Letter

    2) Non-Response Follow-up Letter_ No Form Included

    3) Non-Response Follow-up Letter_ Form Included

C. Selected Screenshots from Electronic Instrument