

SUPPORTING STATEMENT – PART B
U.S. Department of Commerce
U.S. Census Bureau
Business R&D and Innovation Survey
(Forms BRDI-1 and BRDI-1A)
OMB Control No. 0607-0912

B. Collection of Information Employing Statistical Methods

1. Description of the Universe and Respondent Selection

The sample frame consists of a list of about 2 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, exist primarily to make a profit, 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 43,000 companies will be selected annually during the clearance period to represent the approximately 2 million companies on the sample frame. Based on the results of the cognitive testing of the draft questionnaires and instructions, we do not expect a large drop in the overall response rate for the 2011 Survey. For the 2009 Survey, the overall response rate was about 74%.

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 Compustat and Google Finance (4) R&D information from the Bureau of Economic Analysis, and (5) 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 2 million companies is first partitioned into two groups. Group 1 consists of companies where the

status (yes/no) of R&D is known from the pieces of R&D information given above. Group 2 consists of companies where the status of R&D is unknown. Table 1 gives approximate sample frame counts by group based on information from the 2010 Survey.

Table 1 – Approximate Sample Frame Counts by Group

Group/Category	Number of Company Records
Group 1 – Known R&D status	92,000
Group 2 – Unknown R&D status	1,921,000
TOTAL	2,013,000

After the two groups are formed, the sample frame is stratified and companies to be selected with certainty are identified. Group 1 is split into two groups based on the presence or absence of R&D, and the known positive R&D group is stratified by industry category that reflects the detail expected for the published estimates. Group 2 is similarly stratified by industry category. An independent probability sample is then selected within each group and stratum, using constraints on the coefficient of variation achieved on estimated totals to allocate the target sample size of about 43,000 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

Follow-up procedures - Form BRDI-1 companies will continue to have 60 days to report. Reminder letters will be sent to companies that have not responded by mid-May, unless they have been granted extensions. Follow-up letters will be sent in late May, June, and July. The first follow-up package sent in late May will 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 rather than send the second and third notice. These companies account for as much as 85 percent of the value of the data and their responses are critical for the completeness of the estimates. Form

BRDI-1A companies that do not respond within 30 days are sent follow-up letters in April, May, June, and a phone follow-up in July (or until a response is received). Each mail follow-up package includes a duplicate form.¹

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.

Survey form redesign for improved response – Both survey questionnaires and the accompanying instructions were previously redesigned in 2008. Results from the 2008 – 2010 survey cycles as well as cognitive interviews conducted with respondents will be utilized to improve the 2011 questionnaire.

4. Tests of Procedures or Methods

NSF and the Census Bureau plan further improvements of the survey questionnaires and instructions, and other methodological research and improvements to the Survey. Any changes made to the questionnaires will be cognitively tested with respondents prior to implementation.

5. Contacts for Statistical Aspects and Data Collection

Persons responsible for sample design and selection:

Paul L. Hsen, Assistant Division Chief
 Research and Methodology
 Manufacturing and Construction Division
 U.S. Census Bureau
 (301) 763-4586
 (301) 763-7783 (FAX)
paul.l.hsen@census.gov

David L. Kinyon, Chief
 Manufacturing Programs Methodology Branch
 Manufacturing and Construction Division
 U.S. Census Bureau
 (301) 763-7209
 (301) 763-2783 (FAX)
david.l.kinyon@census.gov

Jock Black, Mathematical Statistician
Division of Science Resources Statistics
National Science Foundation
(703) 292-7802
(703) 292-9092 (FAX)
jblack@nsf.gov

Person responsible for data collection:

Mary C. Potter
Assistant Division Chief
Manufacturing & Innovation Indicators
Manufacturing & Construction Division
U.S. Census Bureau
Office: 301-763-4207
mary.catherine.potter@census.gov

Richard S. Hough, Chief
Research, Development, and Innovation Surveys Branch
Manufacturing and Construction Division
U.S. Census Bureau
(301) 763-4823
(301) 763-4718 (FAX)
richard.s.hough@census.gov

Person responsible for analysis of the statistics and publication:

Raymond M. Wolfe, Economist
Research and Development Statistics Program
Division of Science Resources Statistics
National Science Foundation
(703) 292-7789
(703) 292-9091 (FAX)
rwolfe@nsf.gov

Attachments

- A. Cover Letter
- B. Survey Questionnaire (Draft 2011 Form BRDI-1)
- C. Draft Instructions for the 2011 Form BRDI-1/1A
- D. Survey Questionnaire (Draft 2011 Form BRDI-1A)

