A GUIDE TO R & D DATA AT THE CENTER FOR ECONOMIC STUDIES U.S. BUREAU OF THE CENSUS

By

James D. Adams
Center for Economic Studies and University of Florida
and
SuZanne Peck
Center for Economic Studies

CES 94-9 August 1994

information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Room 1587, FB 3, U.S. Bureau of the Census, Washington, DC 20233-6300, (301-763-2065) or INTERNET address snguyen@info.census.gov.

Abstract

The National Science Foundation R&D Survey is an annual survey of firms' research and development expenditures. The survey covers 3000 firms reporting positive R&D. This paper provides a description of the R&D data available at the Center for Economic Studies (CES).

The most basic data series available contains the original survey R&D data. It covers the years 1972-92. The remaining two series, although derived from the original files, specialize in particular items. The Mandatory Series contains required survey items for the years 1973-88. Items reported at firms' discretion are in the Voluntary Series, which covers the years 1974-89. Both of the derived series incorporate flags that track quality of the data. Both also include corrections to the data based on original hard copy survey evidence stored at CES.

In addition to describing each dataset, we offer suggestions to researchers wishing to use the R&D data in exploring various economic issues. We report selected response rates, discuss the survey design, and provide hints on how to use the data.

Keywords: Research and Development, survey data, survey methodology

* The authors wish to thank Ronald Jarmin and Arnold Reznek for helpful comments. Rebecca Turner and Jennifer Cuppy provided excellent typing skills.

TABLE OF CONTENTS

I.Introduction

Expenditures of U.S. corporations on industrial Research and Development (R&D) exceeded $70 billion in 1989.[1] This amount represented 3 percent of the performing firms' sales and once combined with $31 billion of federal R&D, equalled about 2 percent of 1989 GDP. The importance of R&D investments on firm growth and competitiveness has led to many studies of the effects of R&D. The National Science Foundation (NSF) R&D data files available at the Center for Economic Studies (CES) of the U.S. Bureau of the Census (Census) are the principal source of firm level information on R&D. This paper provides a description of the R&D data available at CES as well as offering suggestions to researchers wishing to use the data in exploring various economic issues.

CES has supported a large amount of research based on these R&D data. In Griliches (1980, 1986) he presents his research on firm productivity and R&D expenditures, and his early use of the firm level NSF R&D data at Census. Other studies include the relationship between R&D, total factor productivity, and takeovers, by Lichtenberg and Siegel (1990a, 1990b); the R&D response to import competition by Scherer and Huh (1992); the within firm R&D on productivity by Adams and Jaffe (1994); and the effect of leveraged buyouts on the propensity to perform R&D

---

[1] Research and development includes basic and applied research in the sciences and in engineering, and design and development of prototype products and processes. (NSF, 1992)

by Long and Ravenscraft (1993a, 1993b). These studies have led to ongoing improvement in the quality of the data. At the same time, interest in other research that uses more finely structured questions in the R&D surveys has risen sharply, and accordingly, work on improving the data has recently turned in this direction.

Three major R&D series are held by CES. All three are stored as SAS® data sets.[2] Each consists of data collected from the Survey of Research and Development in Industry. The most basic series transcribes the R&D data held by Census' Industry Division directly into the SAS language and covers the year 1972-92.[3] Henceforth we refer to this as the Transcribed Series. The remaining two series, although derived from the Transcribed files, specialize in particular items. The Mandatory Series contains required survey items for the years 1973-88. Items reported at firms' discretion are in the Voluntary Series, which covers the years 1974-89. Both of the derived series incorporate flags that track data quality. Both also include corrections to data records based on original "hard copy" survey evidence stored at CES. The flags are useful since the quality of survey responses varies widely. The value of corrections to the records

[2]SAS® is the trademark of the SAS Institute.

[3]The 1972-92 transcribed series' data exists at CES as of April 1994. Acquisition of the next year's R&D data usually occurs in the spring two years after the year covered by the survey. For instance, the 1993 R&D data will be available in the spring of 1995.

is debatable, although changes were limited to data that were clearly at odds with hard copy R&D forms. Because, judgement is involved in overwriting records, we have retained the original electronic data files.

The R&D survey is conducted at the firm level and covers approximately 3000 firms reporting positive research dollars.[4] To economize on scarce survey resources at Census as well as limit reporting burdens imposed on survey respondents, the data collection occurs on a mixed annual-biannual basis in the case of large firms with over $1 million in R&D expenditures and every 5-6 years in the case of firms with less than the $1 million threshold. In light of these aspects of R&D survey design and other dataset anomalies, we offer suggestions to researchers wishing to use the R&D data in exploring various economic issues. For example, we report response rates, discuss the survey design, and give hints on how to match the R&D firm level observations to the plant level production data in the Longitudinal Research Database (LRD).

The paper is arranged as follows. Section II provides an overview of the survey design of the R&D survey. Section III discusses the Transcribed series, which contains all the data in

[4] Firms reporting positive R&D are a subset of a much larger sample, most of which report no R&D; the latter set are dropped from the electronic files. However, starting with the 1992 files CES will obtain all firms in the file including those that report no R&D expenditures.

original electronic data files. Section IV describes the Mandatory series of required items and their edit flags. Section V discusses the Voluntary Series consisting of applied R&D by product field, total R&D by state, and basic research by field of science and data quality flags. Section VI offers suggestions to potential users. Section VII concludes.

II. Survey Design

The R&D survey has been conducted annually since 1957. The modern survey largely follows the original design. The design helps NSF meet its legal obligation to produce aggregate time series data on basic and applied R&D by industry, and on company-financed versus federally-financed R&D by industry. The structure of the survey also emphasizes coverage of the maximum dollar value of R&D subject to an upper limit on the number of firms. At the same time, it seeks to limit the amount of information requested each year. The survey design, including the sample selection, the questionnaire design, and the timing of responses to various items, affects the availability of micro data for research in ways that we describe below.

Expenditures included in the R&D totals are described in the instructions to the survey. As described in the instructions, R&D expenditures:

"include all costs incurred to support R&D including R&D depreciation and overhead but excluding capital expenditures. If you perform R&D for others on contract, report the total charged for the work performed including the profit. Include R&D work of consultants performed at company locations. Include R&D performed within company on nonFederal contracts.

Include in R&D expenditures the full cost of all R&D performed. Do not net your R&D expenditures by the amount of royalties received from either non-company organizations or company units, or credits received for R&D work charged or "sold" to other units of the reporting company or to outside organizations.

The relevant costs for R&D usually include but are not limited to the elements listed below:

1. Wages, salaries, and related costs; materials and supplies consumed (or purchased, if consumption figures are not available); costs of computer software used in R&D activities, utilities...; books and periodicals; travel costs and professional dues.

2.Property taxes and other taxes (except income taxes) incurred on account of the R&D organization or the facilities which the R&D organization uses; insurance expense; maintenance and repair, including the maintenance of buildings and grounds; depreciation on buildings, equipment and vehicles; and rentals, if facilities are leased.

3.Company overhead. Estimate a fair share of the cost of any functions which support R&D activities....Items normally covered in overhead include the following: personnel; accounting; procurement and inventory...; other services, including legal, public relations; and salaries and related costs or research executives not on the payroll of the R&D organization.

Exclude R&D contracted out and R&D performed abroad.... Exclude capital expenditures, royalties paid, patent expense, income taxes, and interest; the portion of company-held R&D contracts which were subcontracted to R&D organizations outside the reporting company; and income from the sale of products manufactured in the R&D organization if these were sold to bonafide customers. Also exclude the cost of R&D performed for the company by noncompany organizations of any kind....Exclude fellowships, grants, and gifts to promote R&D." (NSF, 1989)

The R&D sample is chosen from a group of firms known to perform federally-financed R&D and in industries with a concentration of R&D activity. New samples were drawn in 1971, 1976, 1981, and 1987. Appendix 3 provides a detailed description of the methods for compiling the universe from which the 1987 sample is drawn. NSF (1990) provides a more detailed description of selecting the 1987 sample. Appendix 2 includes descriptions of sample designs for survey years 1976-80 and 1981-86. Changes over the years have focused on identifying the universe of firms performing R&D, however, the basic sample selection process has remained the same.[5]

Since the mandate is one of collecting the most industry data from the fewest number of firms, Census samples with 100 percent certainty those firms that are principal R&D performers in each industry. It is believed that these large firms or '4001' firms, so called from the number on the survey form represent 93 percent of R&D expenditures in 1989. Most '4001' firms remain in subsequent sample panels due to their level of R&D expenditures. Census samples smaller R&D performing firms or

[5] In 1992 the sampling methodology for the survey changed. The same basic structure exists as described in this section except now the sample is reselected every year. Census personnel are constantly reviewing sources of R&D information for companies with over $1 million in R&D expenditures. One result of these efforts is an expanded and up-to-date sample with more companies representing firms outside of manufacturing including for-profit R&D labs and computer software developers.

'4002' firms with progressively lower probabilities proportionate to the firm's R&D expenditures. Small firms usually do not participate in adjacent panels. Firms with first time R&D expenditures are not added to the sample as "births" in intervening years of a sample, but only in sampling years. As a result the survey fails to report new emerging R&D efforts for firms not already included in the sample.[6]

The survey design affects the availability of the micro data by limiting how often firms are surveyed. Large '4001' firms are surveyed annually. Small R&D performing firms are surveyed only when a new sample is drawn. As a result, reported data on small R&D performers are limited to three years of CES's R&D databases, 1976, 1981 and 1987. In other years, data on '4002' firms are imputed based on the initial value of R&D and on the average industry growth rate.[7]

Two types of questions exist on the survey forms. Four mandatory data items, domestic net sales, domestic net employment, total R&D, and federally-financed R&D, require firm response. All firms, large and small, are asked these questions when surveyed. The voluntary data are collected less frequently

[6] Starting in 1992 the R&D universe will be resampled every year. For those firms reaching the $1 million threshold of R&D expenditures will be included in the following year's sample as "certainties."

[7] See Long and Ravenscraft (1993b) for a discussion of and tables showing the number of '4001' and '4002' firms with reported and imputed data for 1973-84.

from the large firms. These data include distributions by: applied R&D by product field,[8] total R&D by state, basic R&D by field of science, energy R&D by energy sources, and pollution abatement R&D by form of pollutant. Table 1 lists additional voluntary data items requested. While the mandatory questions have remained in the survey in all years, voluntary questions have been added and subtracted over the 1972-92 period as shown in table 1.

As a result of the survey design, the individual firm records in the electronic files contain varying fractions of reported data, as opposed to missing or imputed data. The survey design determines which firms report in a given year and how much information is asked of that firm. Large firms are not required to respond to voluntary questions.[9] After 1977, detailed voluntary data are only collected in odd numbered years. Large firm records have detailed voluntary data in odd numbered years if they "volunteer" to answer the questions. Small R&D performers (about 60 percent of all firms) are mailed a survey form that consists entirely of mandatory questions. This implies

---

[8] Product fields represent end products of firms. These are grouped in broad categories similar to 2-digit U.S. Standard Industrial Classification codes (Executive Office of the President, 1987). See NSF (1992) for a complete list of product fields.

[9] The R&D survey's "unit-nonresponse [entire firm not reporting] rate ranges from 5 percent to 15 percent. Its item-nonresponse ranges from 1 percent for certain key variables to 10 percent." (King and Kornbau, 1994)

that data observations for small firms contain few reported data items. In addition, responses to all questions are contingent on a survey form having been sent. But we have seen that forms are only mailed to small firms when a new sample is drawn. In other years their data are imputed.

For all these reasons we tend to see relatively few observations containing a lot of detailed R&D data. Nevertheless, sections IV and V of this paper show that the majority of large firms' observations do contain some reported data. Therefore, researchers wishing to undertake a microdata analysis of large R&D performers, generally firms with over $1 million[10] in R&D spending, should be able to carry out their research.

After data collection is complete, the industry series statistics are derived through weighted aggregation of the individual firm data in the sample by primary industry of the firms, using the inverse of the sampling probability as weights.[11] These final statistics are sent to NSF for their approval and publication.

---

[10] The threshold in R&D expenditures for certainty firms prior to 1981 sample was $500,000.

[11] Primary industry of the firm is defined by the largest value of the firm's sales, employment, or payroll, depending on the year, among all industries where it maintains a presence. See appendices 2 and 3 for a discussion of firms' industry assignments across various samples.

III. The Transcribed Series of R&D Files

The most basic R&D files at CES are transcriptions of the mainframe files held by Census' Industry Division into SAS® datasets. The mainframe files are the same files that Census and NSF use to construct the time series of R&D by industry described in section II.[12] The Transcribed Series spans the period 1972-92 at this time. The series is entitled RD72.SSD01-RD92.SSD01. All items in the transcribed files are labelled so that users can display detailed contents of the files by simply running PROC CONTENTS in SAS®. The same is true for the files described in sections IV and V below.

The data items in the Transcribed files come directly from the survey form. Recall that a form that consists of mandatory items is issued to small firms in survey years. The short survey form is sent to large firms in even numbered years (since 1977). In other years a long form is issued to large firms: annually from 1972 through 1977, and in odd numbered years since 1977.

The variables in the transcribed files match those on the long form through 1977. After 1977, they match the variables on the short form in even years and on the long form in odd years. To clarify the array of items in the various years of the R&D survey a facsimile of the short and long forms are included in

[12]The Transcribed Series is the result of work by Steve Andrews and SuZanne Peck at CES.

appendix 5. These forms have changed very little since their initial design. Table 1 provides a simplified outline of the file contents. Appendix 1 provides definitions of basic variables. As seen in the table variables have been added and deleted over the years.

Imputation and last-year-reported flags also appear in the transcribed files from 1983 to present. The imputation flags show an 'R' for reported data or an 'I' for Census imputed data. The Mandatory Series as discussed in section IV provides imputation flags for selected variables for 1973 to 1984. The survey design only calls for imputation of mandatory items in all years. The voluntary items are not imputed at all or are imputed only if prior year reported data exists. NSF has decreased the imputation of data items over the years. Tables in the R&D publications contain variable imputation rates. The last-year-reported flags indicate the last year a firm responded to a particular question.

The firm IDs in the transcribed files differ in form from those in the Mandatory and Voluntary series. The firm IDs identify firm ownership. Under the Census numbering system multi-establishment firms are assigned 10 digit IDs that consist of a unique 6 digit alpha number, beginning with 1 or higher, followed by 4 arbitrary digits. In the transcribed files these final 4 digits vary across years for the same firm. In the other

series these 4-digits are the same across years. This poses a problem when matching observations across years.

Another feature unique to the transcribed files is that they sometimes contain multiple records for one firm. The other data series have summarized the data to the company level (by alpha). The multiple records come from some firms submitting multiple R&D survey forms for their different divisions. Firms may chose to submit their data on one or more forms. Some years may even contain division records and a company summary record. A company summary record ID generally ends in '0000'. Users of the transcribed files have to address these problems of varying final four digits of the ID and multiple reports for one firm, presumably in a similar fashion to the derived files.

IV. The Mandatory Series of R&D Files

This is a brief discussion of a series of files that concentrate on mandatory items in the R&D survey. Table 2 lists the contents of these files, which cover the period 1973-88, and are entitled FLAG73.SSD01-FLAG88.SSD01 (hereafter the FLAG files). The FLAG files contain data amended by edit checks with the original hard copy questionnaires. In addition, they contain imputation flags for years 1973-84 and flags that indicate the availability of firm responses. The data cover three mandatorily reported data items: domestic net sales, domestic net

employment,and total R&D expenditures. This section highlights the variables one may use from these files. Long and Ravenscraft (1993b) include detailed description of the generation of the variables listed in table 2.

One useful variable is SURCODE which indicates the availability of reported data for a firm in a given year. This variable equals '1' if the firm is a '4001' company, a large R&D performing company, and equals '2' if the firm is a small '4002' company. Recall from our earlier discussion that '4001' firms are surveyed every year regarding mandatory R&D data and selected voluntary questions, and every other year concerning their voluntary data (since 1977). Also recall that the '4002' firms are only surveyed in new sample years (1971, 1976, 1981, and 1987). When selecting a year of interest the SURCODE provides an indicator of data availability for large and small firms.

Another very useful feature of the FLAG files is the impute flags for total R&D (FTORD), domestic net sales (FTOSLS), and domestic net employment (FTOEMP) from 1973-88. Prior to 1983 these are the only impute flags that exist for these variables. Flags exist in the transcribed files from 1983 to present.

The edits and imputation flags contained in the FLAG files are essential to future work that uses the R&D data. Table 3 uses the information in the FLAG files to report counts of the average number of firms per year, the average number of firms without imputes of mandatory data per year, and the average

number of large '4001' firms year by varying length of panel. The last column shows an upper bound on the number responding to voluntary questions in the survey. From the row for statistics on '4001' firms the table shows that about half of the 2900 firms surveyed in any given year received survey forms with voluntary data items. For research this is important because these are the largest R&D concerns in the sample.

Table 4 shows the number of firm observations available if a balanced panel is desired. For the years 1972 to 1988 there are 586 firms with continuous data. This number drops to 154 if all firms with imputed mandatory data are deleted. Similarly, the number of '4001' large companies available across 1972 to 1988 is 402. This number also represents an upper limit on the number of firms with voluntary data reported over time. There are 280 large firms with at most one year of imputed data. When considering research using a balanced panel researchers should refer to table 4.

The firm ID number in the FLAG files is easier to use than the firm ID in the Transcribed files. The FLAG files impose a uniform numbering system over time by adding 4 zeroes to the 6 digit multi-establishment alpha numbers in all years. Owing to this uniform system matching establishments over time is easier to do using the IDs in the FLAG files, since the multi-establishment IDs stay the same in all years as long as the firm

is under the same ownership.[13]

Single establishment firm IDs remain unchanged. The IDs for 1972 and 1973 are exceptions. In these years, the final 2 digits of the ID were truncated in the transcribed files which destroys the uniqueness of the ID number for single establishment firms. The truncation requires the use of adjacent year firm information, such as total employment and total net sales, to once more identify the firms.[14] This truncation problem does not affect the uniqueness of the multi-establishment firm IDs.

V. The Voluntary Series of R&D Files

The final series of R&D data emphasizes the distributions of applied R&D by product field, total R&D by state, and basic research by field of science.[15] The data are stored as SAS files named ULTRD74.SSD01-ULTRD89.SSD01 (hereafter the ULTRD files). The years covered are 1974-77, and odd numbered years thereafter. Firm ID numbers in the ULTRD files are of the same structure as

[13] We followed exactly this system in creating the series of voluntary data, ULTRD74.SSD01-ULTRD89.SSD01. Thus, the two edited series are readily combined.

[14] Few firm names and no firm addresses appear in the transcribed files for 1972 and 1973.

[15] The Voluntary Series was created by James D. Adams with the assistance of Jennifer P. Cuppy.

those in the FLAG files and may be matched easily with those in the FLAG files.

The voluntary data in the ULTRD files were checked for errors using hard copy responses of firms and corrected in flagrantly erroneous cases. In addition to these amendments, the files include data quality flags for the voluntary data, which we discuss later in this section. We use the flags to quantify degrees of response to the voluntary questions.

We discuss the contents of the files in terms of broad categories, since roughly 150 variables are involved. The variables are electronically labeled in the files. Table 5 contains a list of all the variables in the ULTRD files.

Most variables described in the tables are the same as those in transcribed files except for the set of data quality flags, which are unique to the ULTRD files. Table 6 is a briefing table on the flags and their codes.[16] All flags are 0-1 dummy variables. In the case of response flags listed at the top of the table (ITRD, IGEOG, IBASIC, IAPPLIED) a value of 1 means that some kind of response, whether real or imputed, appeared for a given R&D distribution, whereas a value of 0 means that there was no response. In this case, a value of 1 means a favorable value.

[16] Flags are missing in 1981 because the 1981 survey forms from which they are derived were destroyed. The following variables are unavailable in the 1981 ULTRD files: TFLAG, GFLAG, BFLAG, TFLAGA, AFLAGA, BFLAGA, LFLAG, PRFLAG, COFLAG, CFFLAG.

In all other cases a value of 1 is an unfavorable value, since it means that the data are lacking in some fashion. In all of these cases in table 6, a value of 1 means that the data are imputed, discrepant in terms of adding up, or else that a given R&D total is not distributed completely among its components.

Tables 7 and 8 provide statistics generated from use of the flags. Table 7 counts the number of responses, whether real or imputed, to questions concerning the distribution of basic research among the sciences, of applied R&D among product fields and of distribution of R&D by state. It does so by summing the number of cases for which IBASIC, IAPPLIED an IGEOG equal 1. We see that the number of basic research responses ranges usually from 200-300 but that a low point is reached in 1987. Likewise we see that the number of applied R&D responses typically varies from 700-1100, but that a minimum is again reached in 1987, when only 500 firms provided information about the distribution of their applied R&D. We see the number of firms distributing data by state goes from 470-2451 with most years around 1000 observations. Notice that the tendency towards declining response is reversed in 1989 for both basic and applied research.

In table 8 we extend the analysis of applied R&D by constructing two flags from the information of the flags listed in table 6. ARESP equals 1 if there is a response where IAPPLIED equals 1, and if the data are not imputed where AFLAG equals 0. AQUAL is more stringent than ARESP. It equals 1 if ARESP equals

1 and if applied R&D by product field sums to a number that is close to total applied R&D where AWIDE equals 0; it equals 0 otherwise. Table 8 reports response rates in columns 2 and 4 with the numerators and denominators of the response rates in columns 3 and 5. The numerator is the number responding and the denominator is the total number of large '4001' firms as indicated by the surcode. The second number bounds the number of respondents, since this is the total number of firms receiving the long form which asks firms to distribute R&D by applied product field.

From columns 2 and 4 in table 8 we see that ARESP is indeed less stringent than AQUAL. We find evidence in the table of declining response, so that by 1987 only a fourth of the firms really answered the voluntary questions about the distribution of applied R&D among product fields.

Tables 7 and 8 show how the flags listed in table 6 can be used to choose observations of good quality from the voluntary data. Tables 9 and 10 apply the exercise carried out in table 8 to the distributions of basic R&D by field of science and the distribution of total R&D by state. The time pattern of the findings here are broadly similar to table 8; response rates decline sharply through 1987, recovering somewhat in 1989. As one would expect, however, the rate of response to basic research questions as in table 9 is quite low, reflecting the meager involvement of industry in this area of research. The response

rate to geographically distributed R&D as in table 10 is generally higher than the applied R&D response rate, perhaps because the whereabouts of R&D are more easily known to respondents.

VI. Some Suggestions for Using the Census R&D Data

In this section we offer a few suggestions for using the data. We begin with suggestions for using the R&D data proper, and conclude with suggestions for merging the R&D with other data sources.

When considering a research project with the R&D data a major consideration is the availability and quality of the data. The R&D publications from NSF prove useful in assessing the data. Many tables in the publications contain one or both of the following symbols:

> "(D), which is used to indicate data withheld to avoid possible disclosure of information about operations of individual companies. This occurs when a small number of companies, usually one or two, accounts for a large percentage of the R&D funds or of scientists and engineers in a particular data cell. Publication of data showing Federal R&D support to companies in R&D-performing industries is most often affected by this rule; and,
>
> (S), which is used to indicate that the imputation rate--the percentage of the statistic estimated by Census staff--exceeds 50 percent. This means respondents failed to provide data for that item on the questionnaire." (NSF, 1989)

A 'D' or 'S' in a table indicates a lack of or poor data for those variables. When considering a research topic a researcher could possibly have difficulty in passing the Census' data disclosure or data quality restrictions if the research is based on variables with many omitted entries in published tables.

Another point to keep in mind is that the transcribed files and the derived files use slightly different firm ID numbers. We have explained how the IDs differ in sections III, IV, and V. We recommend that the IDs of multi-establishment firms in the transcribed files be rendered compatible to the IDs in the derived series by (1), issuing a SAS ATTRIB statement that defines the corrected IDs to have length 10, and their type to be character; (2) using the SAS SUBSTR command to select the first 6 digits of the old IDs of multi-establishment firms, namely the alpha number; and (3) using the concatenation operator || to append 4 zeroes. IDs of single establishments should remain unchanged.

We now turn to the subject of merging the R&D data with the other data available at CES. The first is the merge of R&D data with the LRD production data. In carrying out such a merger it is useful to remember that the R&D data are firm level observations, while the LRD are plant level observations. The match for multi-establishment firms is done on the six-digit alpha numbers. For single units the match must be done with the ten-digit ID number.

One difficulty in performing this match is that the firm ID numbers in the LRD are annually updated to reflect ownership changes, while the R&D firm ID numbers remain the same over the sample period. This means that the user should keep track of nonmatches, noting when they occur, which may signify an ownership change recorded in the LRD. It is a good practice in carrying out the detection of nonmatches to maintain separately named IDs from each of the two data bases. A third commonly named ID can then be used in the matching process itself. This practice prevents the overwriting of either ID. Once the nonmatches are detected the longitudinal SAS file entitled LONGESTB.SSD01 can be used to construct a file of ID changes, so that nonmatches can be retrieved, thereby expanding the set of matches.[17]

Long and Ravenscraft (1993c) discuss matching the R&D data to the Quarterly Financial Report (QFR) data. The QFR is a survey of firms about their "income and retained earnings, balance sheets, and related financial and operating ratios for all manufacturing, mining and trade corporations" (U.S. Bureau of the Census, 1994). The firm identifier numbers are not the same for the QFR and R&D datasets so they cannot be used to

[17] James Adams found that the volume of nonmatches is low. He suggests tracking down only the large firms in the nonmatch file. Work progresses at CES on creating a database of ownership changes over time.

match. Long and Ravenscraft use firm name to link the two databases.

A second firm level dataset of interest is the "Large Company" file which is collected under the "Enterprise Statistics" program.[18] This file contains survey data on firms with 500 or more employees. The data items include firm level capital expenditures, total employment and advertising expenditures. For 1992 the survey includes for the first time a question about total R&D expenditures by the firm. Linking of the two datasets may be performed as with the LRD by alpha for multi-establishment companies and by the ID for single establishment companies.

A source of additional R&D expenditure information available at Census is the Auxiliary Establishment data.[19] These data include statistics on establishments that provide a service to one or more production establishments. Among the list of auxiliary types is research and development labs. Appendix 6 includes a comparison of the two sets of R&D expenditures.

VII. Conclusion

[18] For description of the large company data see U.S. Bureau of the Census (1990b).

[19] For a description of the auxiliary data see U.S. Bureau of the Census (1990a).

This paper described a new resource for the analysis of technical change, the Census-NSF longitudinal R&D database. It has been the result of many years' work by many individuals at U.S Bureau of the Census and elsewhere, but we believe that the new data product will yield valuable research findings concerning the nature of learning in industry that will amply repay the investments in its initial creation. We welcome comments and suggestions for improvement to the R&D database or to this documentation.

REFERENCES

Adams, J.D. and A.B. Jaffe. 1994. "The Span of Effect of R&D in the Firm and Industry." manuscript Center for Economic Studies, Bureau of the Census, Washington, D.C.: May.

Andrews, S.H. and D. Siegel. 1989. "Final Report: Examining the Company-Level R&D Database," manuscript, Center for Economic Studies, Bureau of the Census, Washington, D.C..

Executive Office of the President, Office of Management and Budget. 1987. 1987 Standard Industrial Classification Manual, Washington, D.C.: U.S. Government Printing Office.

_____. 1972. 1972 Standard Industrial Classification Manual, Washington, D.C.: U.S. Government Printing Office.

Griliches, Z. 1986. "Productivity, R&D and Basic Research at the Firm Level in the 1970's," American Economic Review 76 (March): 141-154.

_____ . 1980. "Returns to Research and Development Expenditures in the Private Sector," in J.W. Kendrick and B. Vaccara eds., New Developments in Productivity Measurement, NBER Studies in Income and Wealth No. 44, Chicago: University of Chicago Press, 419-54.

Huh, K. and F.M. Scherer. 1992. "R&D Reactions to High-Technology Import Competition." Review of Economics and Statistics 74 (May): 202-12.

King, C. and M. Kornbau. 1994. "Inventory of Economic Area Statistical Practices, Phase 2: Editing, Imputation, Estimation, and Variance Estimation." ESMD Report Series ESMD-9401, U.S. Bureau of the Census, Washington, DC.

Lichtenberg, F. and D. Siegel. 1990a. "The Impact of R&D Investment on Productivity: New Evidence Using Linked R&D-LRD Data." Economic Inquiry 29 (Winter): 2-13.

____ . 1990b. "The Effect of Takeovers on the Employment and Wages of Central Office and Other Personnel." Journal of Law and Economics 33 (Fall): 383-408.

Long, W.F. and D.J. Ravenscraft. 1993a. "LBOs, Debt, and R&D Intensity." Strategic Management Journal 14: 119-37.

_____. 1993b. "NSF Final Report: The Impact of Corporate Restructuring on Research and Development." manuscript,

TABLE 1
Contents of the Transcribed Series,
RD72.SSD01-RD92.SSD01

Item	**Years Available**
Mandatory Data:	
[1,2]Total R&D	All
[1,2]Domestic Net Sales,	
[1,2]Domestic Net Employment	
Federal-Financed R&D	
Voluntary Data:	All
Company-Financed R&D	
[2]Distribution of total R&D between basic research, applied research, and development	
Number of scientists and engineers	
Distribution of energy R&D by aggregate fuel type	
Total pollution abatement R&D	
Amount of R&D outsourced to foreign firms	
[2]Distribution of R&D costs between wages of research scientists, other labor and materials costs	1972-77, odd numbered years thereafter
[2]Distribution of R&D by state	
[2]Distribution of basic research by field of science	
[2]Distribution of applied R&D by product field	
Distribution of pollution abatement R&D by form of pollution	
Distribution of energy R&D by fuel type	
Amount of R&D outsourced to other domestic firms	
Foreign R&D by Country	1993, odd numbered years
[2]Gestation Lag on R&D	1977 to 1987 in odd
[2]Product versus Process R&D	numbered years
Required by Regulation	1979, 1981, 1983

[1] In Mandatory Series

Table 2
Contents of the Mandatory R&D Series

Variable	Name in File
total research & development	tord
impute flag for tord	ftord
total employment	toemp
impute flag for toemp	ftoemp
total domestic net sales	tosls
impute flag for tosls	ftosls
company name	company
surcode-survey code	rec (=1 if '4001' firm, =2 if '4002' firm)

Table 3
Mandatory Items in the R&D Survey:
Annual Sample Sizes by Length of Panel

Selection Criterion	Years 1972-88	Years 1974-88	Years 1976-88	Years 1978-88	Years 1980-88
All Firms					
N/year	2851	2942	2926	2914	2910
Years	17	15	13	11	9
No imputes					
N/Year	1437	1386	1470	1490	1587
Years	16[a]	15	13	11	9
4001 Firms (Large Companies)					
N/Year	1178	1194	1236	1305	1330
Years	17	15	13	11	9

Notes: [a] Since impute flags begin in 1973, the length of this panel is 1973-88, or 16 years. This table was derived from the Mandatory Data Series.

Table 4
Mandatory Items in the R&D Survey:
Total Sample Size by Length of Panel
for Balanced Panel

Selection Criterion	Years 1972-88	Years 1974-88	Years 1976-88	Years 1978-88	Years 1980-88
All Firms					
N/year	586	820	860	897	910
Years	17	15	13	11	9
No imputes					
N/Year	154	169	213	272	329
Years	16[a]	15	13	11	9
4001 Firms (Large Companies)					
N/Year	402	413	455	617	637
Years	17	15	13	11	9
At most 1 impute					
N/Year	280	323	395	454	513
Years	16[a]	15	13	11	9

Notes: [a] Since impute flags begin in 1973, the length of this panel is 1973-88, or 16 years. This table was derived from the Mandatory Data Series.

Table 5
Principal Contents of the Voluntary R&D Series

Description of Variable Group

distribution of total R&D by state (breakdown is 50 states + District of Columbia in all years)

distribution of applied R&D by product field (breakdown is 32 fields before 1985, 34 fields from 1985 to 1993, 37 fields from 1993 on)[a]

distribution of basic research by field of science breakdown is 11 fields before 1985, 12 fields from 1985 on)[b]

flags for any response, whether real or imputed, to individual items in total, geographic, applied products, and basic research distributions of R&D (ITRD, IGEOG, IAPPLIED, IBASIC)

flags for imputation of any of the individual items in total, geographic, applied products, and basic research distributions of R&D (TFLAG, GFLAG, AFLAG, BFLAG)[c]

flags for imputation of total, applied, and basic R&D (TFLAGA, AFLAGA, BFLAGA)[c,d]

flags for major discrepancies between the sum of the components within total, geographic, applied, and basic R&D distributions and their totals (TWIDE, GWIDE, AWIDE, BWIDE)

flags for 25%+ of totals not allocated to geographic, applied, and basic R&D distributions (GEONA1, APPNA1, and BASNA1)

flags for 50%+ of totals not allocated to geographic, applied, and basic R&D distributions (GEONA2, APPNA2, and BASNA2)

Table 5 (cont)
Principal Contents of the Voluntary R&D Series

Description of Variable Group

dollar amounts not distributed among geographic, applied product, and basic R&D components (GEONDT, APPNDT, BASNDT)

flags for imputation of R&D gestation lags, process vs. product R&D, R&D cost items, company financed vs. federal R&D (LFLAG, PRFLAG, COFLAG, CFFLAG)[c]

applied R&D, basic R&D (ARDT, BT)

mandatory data items (total R&D, domestic net sales, domestic net employment) from both Transcribed Series (RDT, DNS, DNE) and the Mandatory Series (TORD, TOSLS, TOEMP)

miscellaneous identifiers (edited 10 digit company id, 6 digit multi-establishment alpha number, original 10 digit company id, year, surcode)

Notes: [a] In 1985 applied R&D in electrical components and communications equipment was broken up into two separate parts. Applied R&D in optical instruments and in scientific instruments also split into two parts. In 1993, lumber and wood products, paper and allied products leather and software were broken out, and missiles and space vehicles were combined.
[b] NSF added basic research in computer science in 1985.
[c] These flags are missing in 1981 because the 1981 survey forms from which they are derived were destroyed. The flags were done by hand before 1983, but computerized from 1983 onwards.
[d] Available from 1983 to the present.

Table 6
Coding of Data Quality Flags in the Voluntary Series

Type of Flag	Flag Names	Coding
flags for real or imputed response to items in various R&D distributions	ITRD, IGEOG, IBASIC, IAPPLIED	1 if a response to any, 0 if no response
flags for imputation of items in various R&D distributions[a]	TFLAG, GFLAG, AFLAG, BFLAG	1 if any items are imputed, 0 if none are imputed
flags for imputation of total, applied, and basic R&D[a,b]	TFLAGA, AFLAGA, BFLAGA	1 if any items are imputed, 0 if none are imputed
flags for discrepancies between sum of components within R&D distributions and totals	TWIDE, GWIDE, AWIDE, BWIDE	1 if sum differed by a certain percent from the total (usually 10%), or by a dollar amount (often 1 mill. $), 0 otherwise
flags for 25%+ of totals not allocated to various R&D distributions	GEONA1, APPNA1, and BASNA1	1 if 25% or more not allocated, 0 otherwise
flags for 50%+ of totals not allocated to various R&D distributions	GEONA2, APPNA2, and BASNA2	1 if 50% or more not allocated, 0 otherwise

Notes: [a] Not available in 1981. [b] Available beginning in 1983.

Table 7
Number of Real or Imputed Responses to Basic Research, Applied Product Field, and Distributed by State Questions in the R&D Survey
1974-89: Counts of IBASIC, IAPPLIED, IGEOG

Year	Basic Research (IBASIC)	Applied Product Field (IAPPLIED)	Distributed by State (IGEOG)
1974	273	1060	952
1975	245	902	1100
1976	235	906	900
1977	258	1043	1255
1979	266	1079	1221
1981	235	955	930
1983	219	766	470
1985	254	655	2451
1987	173	477	918
1989	311	792	1421

Table 8
True Percents Responding to
Applied Product Field Questions, by year
1974-89

	ARESP		AQUAL	
Year	Percent Responding	Number out of the total	Percent Responding	Number out the total
1974	81.9	865 out of 1056	80.4	849 out of 1056
1975	75.6	638 out of 844	56.3	475 out of 844
1976	76.2	700 out of 919	75.7	696 out of 919
1977	72.9	628 out of 862	72.6	626 out of 862
1979	58.6	710 out of 1211	57.7	699 out of 1211
1983	42.3	558 out of 1320	41.1	453 out of 1320
1985	43.8	592 out of 1352	43.8	592 out of 1352
1987	24.5	420 out of 1714	24.4	419 out of 1714
1989	46.2	746 out of 1613	43.3	698 out of 1613

Notes: Total equals all '4001' firms. ARESP equals 1 if there is a response to the applied product field question and if the data are not imputed; it equals 0 otherwise. AQUAL equals 1 if ARESP equals 1 and if the applied product field data add to total applied R&D.

Table 9
True Percents Responding to
Basic Research Questions,by year
1974-89

	BRESP		BQUAL	
Year	Percent Responding	Number out of the total	Percent Responding	Number out the total
1974	21.9	231 out of 1056	21.3	225 out of 1056
1975	21.4	181 out of 844	15.5	131 out of 844
1976	19.2	176 out of 919	19.2	176 out of 919
1977	19.4	167 out of 862	19.4	167 out of 862
1979	15.4	186 out of 1211	14.9	180 out of 1211
1983	11.6	153 out of 1320	9.8	129 out of 1320
1985	10.2	138 out of 1352	10.2	138 out of 1352
1987	5.5	94 out of 1714	5.5	94 out of 1714
1989	11.5	185 out of 1613	11.1	179 out of 1613

Notes: Total equals all '4001' firms. BRESP equals 1 if there is a response to the basic research questions and if the data are not imputed; it equals 0 otherwise. BQUAL equals 1 if BRESP equals 1 and if the basic research data can be allocated and add to total applied R&D.

Table 10
True Percents Responding to
Geographic R&D Questions,by year
1974-89

	GRESP		GQUAL	
Year	Percent Responding	Number out of the total	Percent Responding	Number out the total
1974	72.4	765 out of 1056	69.3	732 out of 1056
1975	71.0	599 out of 844	68.0	574 out of 844
1976	77.1	709 out of 919	75.0	689 out of 919
1977	78.1	673 out of 862	75.8	653 out of 862
1979	75.4	913 out of 1211	72.6	879 out of 1211
1983	27.1	358 out of 1320	25.4	335 out of 1320
1985	46.7	631 out of 1352	45.2	611 out of 1352
1987	25.6	439 out of 1714	25.0	429 out of 1714
1989	49.3	796 out of 1613	46.7	754 out of 1613

Notes: Total equals all '4001' firms. GRESP equals 1 if there is a response to the geographic R&D questions and if the data are not imputed; it equals 0 otherwise. GQUAL equals 1 if GRESP equals 1 and if the geographic data can be allocated and add to total R&D.

Center for Economic Studies, Bureau of the Census, Washington, DC.

_____. 1993c. "The Quarterly Financial Report (QFR) Database." manuscript, Center for Economic Studies, Bureau of the Census, Washington, DC.

National Science Foundation (NSF). 1992. Research and Development in Industry: 1989. by J.R. Gawalt. NSF 92-307, U.S. Government Printing Office, Washington, DC.

____ . 1990. Estimating Basic and Applied Research and Development in Industry: A Preliminary Review of Survey Procedures. by E.I. Collins. NSF 90-322, U.S. Government Printing Office, Washington, DC.

_____. 1989. Research and Development in Industry: 1987. NSF 89-323, U.S. Government Printing Office, Washington, DC.

_____. 1984. Research and Development in Industry: 1982. NSF 84-325, U.S. Government Printing Office, Washington, DC.

_____. 1981. Research and Development in Industry: 1979. NSF 81-324, U.S. Government Printing Office, Washington, DC.

U.S. Bureau of the Census. 1994. Quarterly Financial Report for Manufacturing, Mining, and trade Corporations Fourth Quarter, 1993. Series QFR-93-4, U.S. Government Printing Office, Washington, DC.

_____. 1990a. 1987 Enterprise Statistics: Auxiliary Establishments. Series ES87-2, U.S. Government Printing Office, Washington, DC.

_____. 1990b. 1987 Enterprise Statistics: Large Companies. Series ES87-1, U.S. Government Printing Office, Washington, DC.

APPENDIX 1
Survey Definitions

[Excerpt from NSF(1989).]

Research and development--Basic and applied research in the sciences and engineering and the design and development of prototypes and processes. This definition excludes quality control, routine product testing, market research, sales promotion, sales service, research in the social sciences or psychology, and other nontechnological activities or routine technical services.

Basic research--Original investigations for the advancement of scientific knowledge not having specific immediate commercial objectives, although such investigations may be in fields of present or potential interest to the reporting company.

Applied research--Investigations directed to the discovery of new scientific knowledge having specific commercial objectives with respect to products or processes. This definition differs from that of basic research chiefly in terms of the objectives of the reporting company.

Development--Technical activities of a nonroutine nature concerned with translating research findings or other scientific knowledge into products or processes. Not included are routine technical services to customers or other activities excluded from the foregoing definition of R&D.

Funds for research and development--Operating expenses incurred by a company in the conduct of R&D in its own laboratories or other company-owned or -operated facilities. These expenses include wages and salaries, materials and supplies consumed, property and other taxes, maintenance and repairs, depreciation, and an appropriate share of overhead, but exclude capital expenditures.

Company-financed research and development--Cost of company-sponsored R&D actually performed within the company. These data therefore do not include the cost of R&D supported by companies but contracted to outside organizations, such as research institutions, universities and colleges, nonprofit organizations, or (to avoid double-counting) other companies. Since it is a survey of R&D performers, industrial firms than undertake R&D supported by other companies, however, do report the funds received in payment for the R&D work the perform. These monies are classified under the industries of the <u>performing</u> companies.

Federally financed research and development--Receipts for work done by the company on Federal R&D contracts or subcontracts and R&D portions of procurement contracts and subcontracts.

Federally funded research and development centers (FFRDCs)-- Organizations administered by industrial, educational, or other institutions on a nonprofit basis; they conduct R&D almost exclusively for use by the Federal Government. R&D expenditures of industry-administered FFRDCs are included in data showing Federal R&D support to industry under the industry classifications of the administering firms.

R&D scientists and engineers--The January number of those engaged full time in R&D and the full-time-equivalent (FTE) of those working part time in R&D. Scientists and engineers are defined as persons engaged in scientific or engineering work at a level that requires knowledge of physical, life, engineering, or mathematical science equivalent at least to that acquired through completion of a 4-year college program with a major in one of those fields.

Employment--Total number of persons domestically employed by R&D-performing companies in all activities during the pay period that includes the 12th of March. These data are not completely comparable with the data on R&D scientists and engineers described in the foregoing paragraph because the earlier data were collected in January of each year.

Net sales and receipts--Recorded dollar values for goods sold or services rendered by R&D-performing companies to customers (outside the company), including the Federal Government, less such items as returns, allowances, freight, charges, and excise taxes. Domestic intracompany transfers and sales by foreign subsidiaries are excluded, but transfers to foreign subsidiaries and export sales to foreign companies are included.

Geographic area covered--Includes those operations located in the 50 States and the District of Columbia. Company-sponsored R&D performed outside the United States by foreign subsidiaries of U.S. domestic companies is reported as one total.[1]

[1] This is true for all years up to 1993. In the 1993 survey companies are asked to breakout their R&D performed overseas by country.

Industry classification--Census Bureau staff assigned a company-level Standard Industrial Classification (SIC)[2] code to each company. For multi-establishment companies, single SIC codes--representing the most dominant economic activity (in terms of total payroll)--were assigned. Data for the following industry groupings [with SIC code(s) shown in parentheses] are published in this report:

- Food and tobacco (20, 21)[3]
- Textiles and apparel (22, 23)
- Lumber, wood products, and furniture (24, 25)
- Paper and allied products (26)
- Chemicals and allied products (28)
  - Industrial chemicals (281-82, 286)[4]
  - Drugs and medicines (283)
  - Other chemicals (284-85, 287-89)[5]
- Petroleum refining (29)
- Stone, clay, and glass products (32)
- Primary metals (33)
  - Ferrous metals and products (331-32, 3398-99)
  - Nonferrous metals and products (333-36)
- Fabricated metal products (34)
- Machinery (35)
  - Office, computing, and accounting machines (357)
  - Other machinery, except electrical (351-56, 358-59)
- Electrical equipment (36)
  - Radio and TV receiving equipment (365)
  - Communication equipment (366)
  - Electronic components (367)
  - Other electrical equipment (3611-64, 369)
- Transportation equipment (37)
  - Motor vehicles and motor vehicles equipment (371)
  - Other transportation equipment (373-75, 379)
  - Aircraft and missiles (372, 376)[6]

---

[2] Executive Office of the President (1987).

[3] Until 1984, tobacco products (SIC 21) was included with "other manufacturing industries".

[4] The classification of "Industrial chemicals" was revised to include SIC Group 286, Industrial organic chemicals.

[5] See footnote number 4.

[6] Companies primarily engaged in the manufacture of ordnance and accessories, including complete guided missiles, are grouped with companies primarily engaged in the manufacture of aircraft

Professional and scientific instruments (38)
Scientific and mechanical measuring instruments (381-82)
Optical, surgical, photographic, and other instruments (383-87)
Other manufacturing industries[7]--printing and publishing (27), leather products (31), and miscellaneous manufacturing industries (39)
Nonmanufacturing industries--forestry (08); mining (10-12, 14); construction (15-17); transportation, communications, electric, gas, and sanitary services (40-49); wholesale and retail trade (50-59); finance, insurance, and real estate (60-67); personal and business services (72-73); health services (806-07); and engineering, accounting, research, management, and related services (87)

Classification of reporting units--The company or corporate family that includes all establishment under common ownership or control is the basic reporting unit. All R&D expenditures and scientists and engineers of each company are classified into a single SIC code and size-category.

---

and parts because of the close similarity of their R&D activities.

[7] See footnote number 3.

APPENDIX 2
Methodology of Survey for 1976-87 Survey Years

[Excerpt from NSF(1981,1984).]

1976-81 Survey Years

The sample used for the 1976-1981 Surveys of Industrial Research and Development represented all manufacturing industries and those nonmanufacturing industries known, on the basis of earlier, more detailed samples, to conduct or to finance research and development. The sampling unit for the survey was the company, defined as a business organization consisting of one or more establishments under common ownership or control. A new panel for the R&D survey is selected approximately every five years. The latest panel was selected for the 1976 survey, the first since the 1971 survey. Approximately 11,500 manufacturing and nonmanufacturing companies are included in the current sample, which consists of about 4,500 certainty companies (those with 100 percent chance of inclusion in the panel) and about 7,000 noncertainty companies.

The basic tool for the survey is Form RD-1, which seeks detailed R&D information from respondents. Companies in the new panel which had received an RD-1 form in the old panel (1971-75) once again received an RD-1 form in 1976 (about 1,100 companies). The remaining certainty and noncertainty companies in the new panel received an RD-2 survey form[1] in 1976. Form RD-2 is an abbreviated version of RD-1 and is only mailed to companies in the year in which a new sample is drawn. The purpose of form RD-2 is to canvass smaller R&D performers with a minimum of reporting burden. Once the RD-2 forms from the survey respondents in 1976 were received and tabulated, they were reviewed for size. Those RD-2 companies which reported R&D expenditures of $500,000 or greater were converted to Form RD-1 reporters and were included with other RD-1 companies in the 1977-79 surveys. There were about 450 such companies. The remaining RD-2 companies were not mailed another form; Census estimated their data based upon their 1976 report.

All manufacturing and selected nonmanufacturing companies (in SIC's 40, 7391-92, 7399, and 8911) with 1,000 or more employees were included in the sample certainty. Manufacturing and selected nonmanufacturing companies with fewer than 1,000 employees were sampled at rates depending upon their industry and employment size. The source of this sample was the 1974 Standard Statistical Establishment List (SSEL). For 1976, the SSEL was used for the first time as a source for the R&D sample. For

[1] [The RD-2 form became the RD-1A form for the 1981 panel.]

other nonmanufacturing industries, the sample was based on the 1966 records of the Social Security Administration.

Each year the Census Bureau reviews the annual lists of R&D contractors published by the Department of Defense (DOD) and National Aeronautics and Space Administration (NASA) to ensure that the large contractors are included in the sample. For the 1979 survey, the R&D-performing manufacturing companies from the 50 largest NASA contractors were included in the reporting panel with certainty.

1981-87 Survey Years

The sample used for the 1981 Survey of Industrial Research and Development represents all manufacturing industries and those nonmanufacturing industries known-on the basis of earlier, more detailed samples-to conduct or to finance research and development. The sampling unit for the survey was the company, defined as a business organization consisting of one or more establishments under common ownership or control. A new panel for the R&D survey is selected approximately every five years. The newest sample was selected for the 1981 survey (the first since the 1976 survey). Approximately 11,500 manufacturing and nonmanufacturing companies are included in the current sample, which consists of about 4,500 certainty companies (those with 100-percent chance of inclusion in the panel) and about 7,000 noncertainty companies.

The basic tool for the survey is form RD-1, which is used to collect detailed R&D information. Companies in the new panel that received an RD-1 form in the old panel once again received an RD-1 form in 1981 (about 1,200 companies). The remaining certainty (about 3,300) companies and noncertainty companies (about 7,000) in the new panel received RD-1A survey forms for 1981. Form RD-1A is an abbreviated version of RD-1 and is only mailed to companies in the year in which a new sample is drawn. The purpose of form RD-1A is to canvass smaller R&D performers with a minimum of reporting burden. Once the RD-1A forms were received and tabulated from the survey respondents in 1981, they were reviewed for total R&D expenditures. Those companies which reported R&D expenditures exceeding $1,000,000 on the RD-1A form were added to the survey panel consisting of those firms that receive the RD-1 form annually. There were about 575 such companies in 1981. The remaining RD-1A companies are not mailed another form but, in subsequent years, data for them are estimated by Census based upon their 1981 reports.

The universal frame from which the latest sample panel was selected was created from two sources-the 1981 Standard Statistical Establishment List (SSEL) for single units and the 1981 Enterprise Statistics Multi-establishment file.

There are about 3.5 to 4 million singleunit firms in the 1981 SSEL file (including nonmanufacturing firms). There are 5.6 million multi-establishment companies in the 1981 Enterprise file, of which 296,146 companies have identified themselves as engaged primarily in manufacturing. All manufacturing industries and selected nonmanufacturing industries (SIC's 49, 7391, 7392, 7399, and 8911) were considered within the scope of the survey.

Companies in these industries with 500 or more employees were included in the panel with certainty. For the companies with fewer than 500 employees, a measure of size was assigned to each company based on an estimate of its total R&D expenditures. Probabilities of selection were then assigned based on the measure of size. Finally, a sample selection process gave each company an independent chance of being included in the sample.

Each year Census reviews the annual lists of R&D contractors published by the Department of Defense (DOD) and the National Aeronautics and Space Administration (NASA) to ensure that the large contractors are included in the sample. For the 1981 survey, the R&D-performing manufacturing companies from the largest DOD and NASA contractors were included in the reporting panel with certainty.

APPENDIX 3
Methodology of Survey for 1987-present Survey Years

[Excerpt from NSF(1989).]

- The annual Survey of Industrial Research and Development has been conducted for NSF by Census for the past 30 years.

- All companies, both foreign and domestic, that perform R&D in the United States are included or represented.

- All companies that annually spend more than $1 million on R&D in the United States receive a survey form every year.

- Privately held companies are included.

- Respondents are provided detailed definitions to guide them on which expenses to include or exclude from the data they provide.

- Census staff conduct the survey under Title 13 of the U.S. code which prohibits publication or release of data that may reveal information about individual companies.

- It is a company-rather than an establishment-based survey. Therefore, all R&D data for each company are placed with the major Standard Industrial Classification (SIC) code of the firm for all tables, except those showing R&D expenditures by product field.

Introduction

NSF first sponsored a survey of industrial R&D in 1953. Since then, the scope of the survey has gradually been expanded and refined in response to an increasing need for more detailed information on the Nation's R&D effort.

The 1987 survey of industrial R&D is the 31st in the annual series sponsored by NSF and conducted by Census, Department of Commerce. Industry Studies Group staff of NSF's Division of Science Resources Studies monitors the survey.

The primary focus of these data-gathering efforts is on U.S. industry as a performer of, rather than as a source of funds for, R&D. Thus, data on Federal support of R&D activities performed by industry are collected, but data on industry support of R&D undertaken at colleges and universities and other nonprofit organizations are not collected.[1] They are, however, included

[1] Data on industry funding of R&D performed at universities and colleges are collected in the Annual Survey of Scientific and

with the total amount of R&D funds contracted to outside organizations.

The statistics are subject to response and concept errors caused by different respondent interpretations of the definitions of R&D activities provided in the survey instructions and by variations in company accounting procedures. Consequently, the data are better indicators of changes in, rather than absolute levels of, R&D spending and personnel.

Data quality has improved substantially since the first industry R&D survey was undertaken, mainly as a result of respondents' adoption of more accurate and sophisticated accounting procedures. In addition, NSF and Census staff have endeavored to reduce response and concept errors arising from difficulties in interpreting or applying survey definitions.

NSF staff are aware of the increased reporting burden placed on industry from all sources in recent years. To reduce this burden, the detailed questionnaire, which has been in use with slight modifications since the beginning of the survey, is now mailed only biennially, in odd-numbered years; abbreviated forms containing only the most crucial data elements are sent to survey respondents in the intervening, even-numbered years. The shortened survey form was used for the first time to collect industrial R&D data for 1978.

Methodology of Survey[2]

The data available are based on a probability sample, selected and first used for survey year 1987. The universe from which the probability sample, or "panel", was drawn includes companies in all manufacturing industries and a select number of nonmanufacturing industries known, on the basis of earlier samples, to conduct R&D. The sampling unit for this survey is the company, defined as a business organization consisting of one or more establishments under common ownership or control.

The Standard Statistical Establishment List (SSEL), which contains information on 3.5 to 4.0 million establishments (that are either entire companies or parts of companies) was the universe frame used to select the 1987 panel. Establishment-level data were summed, if necessary, to the company-level, and Census staff assigned a single SIC code-the SIC code of the

---

Engineering Expenditures at Universities and Colleges. More information about this survey is available from the Universities and Colleges Studies Group of NSF's Division of Science Resources Studies.

[2] This section was prepared in the Industry Division of the Bureau of the Census.

establishment(s) having the highest dollar-value of payroll-to each company.

Several innovations were introduced into this most recent sample design to improve the quality of the sample vis-a-vis earlier sample designs. (The previous panel was selected and first used for the 1981 survey and was also used in subsequent annual surveys until 1987).

Frame Creation

From the outset in the latest sample selection, the major goal was to eliminate from the frame, to the greatest extent possible, companies unlikely to have R&D programs. This would minimize the number of sampled companies without R&D activity. To accomplish this objective, two steps were taken:

1. NSF staff narrowed the list of "in scope" nonmanufacturing industries by eliminating those known to have a little or no R&D activity. Thus, companies in the eliminated nonmanufacturing industries had no chance of being selected. This gave companies in the remaining nonmanufacturing or in manufacturing industries a greater probability of selection (than in past sample selections).

2. Additional companies--even some in "in-scope" industries--were eliminated from the universe frame because they had fewer than a specified number of employees. An assumption was made that companies with only a small number of employees in some (for the most part nonmanufacturing) industries are unlikely to have R&D activity. Those companies were eliminated from the frame. NSF staff provided an employment cutoff for each industry group.

In another effort to improve coverage of R&D-performing companies, NSF staff provided names of firms that were to be included in the sample with certainty. Most of these companies would have received questionnaires anyway because they met other established criteria; the few that did not were added to the panel.

In addition, Census staff reviewed lists of R&D contractors published by the Department of Defense (DOD) and the National Aeronautics and Space Administration (NASA) to ensure that all large industrial DOD and NASA R&D-performing contractors were included in the panel with certainty. Further, all companies with more than 500 employees in "in-scope" industries were sampled with certainty.

All certainty companies-on lists provided by NSF staff, on lists of DOD and NASA contractors, companies with more than 500

employees, and previous panel members-are self-representing, i.e., they have sampling weight of unity (1.00).

Based on (1) SIC code, (2) total employment cutoffs, (3) inclusion on an NSF, DOD, or NASA list, or (4) previous panel membership, approximately 154,000 companies were identified as "in scope" of the survey and therefore were included in the sampling frame. The effect of the new efforts aimed at improving coverage is demonstrated by a sharp reduction in the size of the total universe; it dropped from about 450,000 companies in the 1981 sampling frame to 154,000 companies in the latest operation.

It is likely that a small number of companies actually engaged in R&D activity were omitted from the frame as a result of these first-time sample selection operations. It was agreed, however, that the benefit derived from the new operations-greater sampling efficiency resulting in improved national estimates of R&D expenditures and employment-far outweighed the cost.

Probability Proportionate to Size

As with most types of economic surveys, the sample selection process for the industrial R&D survey used probabilities proportionate to size (pps). That is, "large" companies have a proportionately higher probability of selection than do "small" companies, where large or small is measured relative to the statistic being estimated.

For the R&D survey, size should be determined by the amount of a company's R&D expenditures. Unfortunately, except for the portion of the universe frame that was in the current panel, it was impossible to know what these R&D expenditure values were. One logical solution was to impute each companies R&D expenditures and base the probability of selection on these imputed values. (The same strategy was employed in the 1981 sampling operation).

Each company was assigned a probability of selection, based on the size of its estimated R&D expenditures. The size of each company's R&D expenditures was estimated by Census using a relationship linking the size of its R&D expenditures to its employment.[3] This relationship was developed for each SIC from data collected in the then most recent (1985) R&D survey. Thus, within each SIC, the larger the number of employees, the higher the probability of selection for inclusion in the sample.

Clearly, this strategy has some weaknesses. Even with refinement of the universe frame, as described in the foregoing paragraph, a large number of companies on the frame have no R&D

[3] Since company employment was known for the universe, it was possible to use this relationship to impute R&D expenditures values for all companies in the frame.

activity. But this procedure treated all companies as if they do. Although they might not have been assigned the most appropriate measure of size and, hence, probability of selection, it is reasonable to assume that large companies are more likely to have R&D programs than small companies (thus giving large companies greater probability of selection) rather than to treat all companies equally. An additional consequence of this assumption is discussed later.

One further adjustment was applied that was not made in previous sample selections. This was based on the assumption that multi-establishment companies of a given size and in a given industry would on average be expected to have more R&D activity than a single-establishment company of the same size and in the same industry. Once again, 1985 panel data were used to develop this adjustment factor. Finally, it should be noted that for companies in the previous panel, their actual reported R&D activity was used in lieu of an imputed value and was unadjusted.

Sample Allocation and Relative Standard Error Constraints

The sampling program utilized for this operation allowed parameters to be assigned permitting the sample to be allocated across various levels or strata that correspond to industry groupings. This procedure permitted a desired sample size or a desired sampling error to be achieved for each strata. Estimated errors of total R&D estimates for these strata were not to exceed certain levels. The only constraint in achieving these results was that the total sample size across all the strata could not exceed 12,000-13,000 companies. (The amount of funds provided by NSF determined the size of the sample to be drawn). NSF staff provided relative rankings for each industry group-high, medium, or low-to determine the precision of the estimate. An actual translation to what high, medium, or low meant specifically could not be determined until Census Staff arbitrarily investigated several sampling error levels, computed what sample size these levels implied, and applied the constraint of the total sample size of 13,000. The result of this investigation led to the following criteria:

a. High precision: sampling error not to exceed 2%
b. Medium precision: sampling error not to exceed 5%
c. Low precision: sampling error not to exceed 10%

Based on the desired precision these criteria suggested a total sample size of approximately 13,500. This number was not excessively beyond the stated limit of 13,000, so this was the sample size parameter decided on for the selection process.

One limitation should be noted. Sampling errors were controlled using a universe total that in large part was

improvised; that is and as noted above, an R&D value was assigned to every frame record, although in reality many companies in the sampling frame have no R&D expenditures. The value was an imputed value for the great majority of companies in the frame. As a consequence, the estimated universe and the distribution of individual company values bore little resemblance to reality. Estimates of sampling variability were nevertheless based on this distribution. The presumption was-and this had been confirmed in the previous sample selection-that actual variation would be less than that estimated because so many of the sampled companies have true R&D values of zero, not the widely varying values that were imputed. Thus, the 2 percent, 5 percent, and 10 percent error levels described above are conservative.

The particular sample selected is one of a large number of samples of the same type and size that, by chance, might have been selected. Estimates from each of the different samples would differ somewhat from each other and from the results of a complete canvass conducted under essentially the same conditions as the survey.

In addition to sampling error, the estimates are subject to nonsampling error that would also occur if a complete canvas were to be conducted under the same conditions.

Sample Selection

The sample selection program was run with a specified sample size (expected) of 13,500 and with other parameters set to assure compliance with the relative standard error constraints. An actual sample of 13,917 was selected. There are two reasons why the actual sample size differs from the specified:

First, the program uses independent sampling. Each company had an independent chance of selection based on its assigned probability; the selection (or nonselection) of a company was completely independent of the selection of any other company.

In independent sampling, sample size is itself a random variable. Theoretically, a sample of size 0 or a sample the size of the entire universe is possible, but the probabilities of these extremes are so small that these are nearly impossible situations. The actual sample size is usually quite close to the specified size. If there is too much deviation, the program is simply executed again.

Second, a minimum probability rule was imposed. As noted earlier, the sampling program assigns probabilities proportionate to size (where size in this case is the imputed R&D value assigned each company). Selected companies that are vastly larger than their assigned values can have adverse effects on the estimates once the data are collected. To lessen these effects, the maximum weight a company can assume was arbitrarily controlled by specifying that the probability of selection cannot

be less than a certain value. If the probability based on it size is less than this minimum value, then it is set equal to this value. The consequence of raising these original probabilities to the minimum probability is to raise the expected sample size. It is likely that most of the difference between the specified sample size and the actual sample size is due to this rule.

The Annual Panel

A panel is a group of companies that receive a survey questionnaire, the RD-1, annually. The following is a description of how the new panel was formed from the sample.

The basic tool for the survey is form RD-1, which is used to collect detailed R&D information. Companies in the new sample that were in the old panel and had received a 1986 RD-1 form (1,095 companies) once again received an RD-1 form for 1987. The remaining certainty (6,903) and noncertainty (5,919) companies in the new sample received an RD-1A survey form[4] for 1987. Form RD-1A is an abbreviated version of RD-1 and is generally mailed to companies only in the year in which a new sample is drawn. The purpose is to canvass, with a minimum of reporting burden, smaller R&D performers.

Of the 13,917 companies that received a form, 3,793 respondents reported that their companies had R&D expenditures. The 3,793 companies were ranked by total R&D (both companies' own and Federal) funds within each SIC code. All companies with over $1 million in total R&D expenditures were placed on the RD-1 panel. In some industries, companies with less than $1 million in R&D expenditures were also added to the panel to ensure 95 percent coverage of the R&D total for each industry. All companies in the panel will receive the RD-1 questionnaire annually until the next sample is drawn. The other RD-1A companies (with less than $1 million in R&D expenditures) will not receive another questionnaire; their data will be estimated, using their 1987 reports, in subsequent years by Census staff.

The RD-1 panel increased from 1,095 companies in 1987 to 1,795 companies in 1988. A few companies report by establishment on more than one form. Accounting for multiple reports from companies, the number of mailing units increased from 1,252 to 1,946 for 1988.

[4] See appendix 5 for a sample of the RD-1 and RD-1A survey forms.

Table A-2[5] contains information, by industry, on the number of companies in the sample having R&D expenditures and the composition of the 1988 RD-1 annual survey panel.

The survey questionnaires were mailed in January 1988, and nonrespondents received followup letters by mail. Since total R&D expenditures, Federal R&D funds, net sales, and employment are included in Census' mandatory statistical program, form MA-121s, which are used to collect these mandatory items, were mailed to the few companies that had not returned form RD-1 for 1987. When companies fail to provide the requested information, the missing data are estimated by using industry averages and several different methodologies that rely on data provided in earlier years.

[5] See NSF (1989) for table A-2 which shows the number of companies in the R&D expenditures universe, sample and panel.

APPENDIX 4
Comparability of Data Over Time

[Excerpt from NSF(1989). This excerpt describes how the published data is generated over time. It provides some useful information on the firm-level data.]

Several procedures are undertaken to maintain the reliability of the industry R&D [published] time series:

Two-Year Comparability

Before mailing the survey forms each year, data reported by respondents the previous year-or two years earlier for items asked only in odd-numbered years-are imprinted on the questionnaires. Respondents are asked to adjust the data for the previous year(s) as necessary to make them comparable to data provided for the current year. Such adjustments are necessitated, for example, by changes in reporting concepts or changes in company structure. Thus, there is comparability in data from the survey over any 2-year period. To maintain consistency, the employment-size classification of any company affected by such changes is adjusted so that the company is tabulated in the same employment-size category for two consecutive years.

These adjustments can be examined by comparing data for the same year reported in two succeeding periods, e.g., 1984 data appearing in the 1984 edition of Research and Development in Industry may differ from 1984 data in this volume. Totals for broad classifications are likely to be very close in the two editions; larger differences are more noticeable in the finer detail. These differences underscore the point that the measures are approximate and indicative rather than precise.

Historical Data Revisions

The industry R&D survey data are revised periodically, usually because of changes in company SIC classifications. Companies may shift from one industry into another because of any of the following: (1) the growth and/or decline of product lines, (2) the merger of two or more companies, (3) the acquisition of one company by another, (4) divestiture, or (5) the formation of conglomerates. If Census Bureau staff are aware of the year in which changes #2, #3, #4, or #5 occurred (respondents are asked about changes in ownership on the questionnaire), data are reclassified in the new industry for the year the change actually occurred. If a change was not discovered until the selection of a new panel or if it could not be determined when a shift actually occurred (i.e., #1), other methodologies were used to

move a company out of one industry and into another. Since 1967, three revisions in the data covering the periods 1967-76, 1976-81, and 1981-87 were made to adjust data of companies that changed industries. These are described below.

The 1967-76 Period

The SIC codes assigned to companies in the panel for the years 1967 through 1975 were based on data reported in the 1967 census. The 1974 SSEL file was used to assign SIC codes to companies in the next panel chosen for the 1976 survey and for revised 1975 data received in the 1976 survey. The SIC codes of companies in the 1967 and 1976 panels were examined to determine which companies had changed classifications. Since it was not known in which year changes actually occurred, data of companies that had changed SIC codes were revised for the years 1968 through 1974 to smooth the changes over the period 1967-76. To illustrate, if a company was originally in SIC A in 1967 but was discovered to be in SIC B in 1974, its data for 1967-74 were allocated between the two industries as follows: 1967--all of the company's data was retained in industry A; 1968--14.3 percent of the company's data was allocated to industry B; and the remainder retained in industry A; 1969--28.6 percent was allocated to industry B and the remainder retained in industry A; and so on until 1974, when all of the company's data was allocated to industry B.

The 1976-81 and 1981-87 Periods

Similar revisions in the industry R&D data were made for companies in the panels drawn in 1976 and 1981 used for the years 1976-80 and 1981-87, respectively, but a different methodology was used.

When the most recent panel (1987) was selected, companies were assigned SIC codes from the SSEL File. Prior-year (1986) data were collected in the 1987 survey.[6] These 1986 data were presumed to be more accurate than those collected in the 1986 survey because they not only reflected updated SIC codes, but also were obtained from a larger panel providing better coverage of U.S. industry. Thus data obtained for 1981-86 using the panel selected in 1981 were revised subject to the following constraints:

---

[6] [In the R&D files at CES a file exists for both the 1986 data collected in 1986 and the 1986 data collected in 1987. The set of firms is different between the two files. The first 1986 file represents the 1981 sample. The next 1986 file reflects the firms surveyed in the 1987 survey.]

1. Data for 1981 (revised from the 1982 survey) would remain unchanged since this was the first year the 1981 panel was used and that panel was an accurate reflection of company SIC codes in that year.

2. Data from 1986 collected in the 1987 survey would be used instead of the 1986 data collected in the 1986 survey.

An algorithm was used to link data from 1981 with those collected in 1987, preserving, to the greatest extent possible, year-to-year trends in data for each industry by revising data for the years 1982, 1983, 1984, and 1985. Interested persons should contact the Census Bureau to obtain further information about the construction and content of the algorithm.

The following data elements were adjusted using the methodologies just described: Funds spent on R&D (total, Federal, and companies' own); number of FTE R&D scientists and engineers; total and company R&D funds as a percent of new sales; cost per R&D scientist or engineer; and basic research expenditures. No adjustments were made in other data elements.

APPENDIX 5
RD-1 Short and Long Form, and RD-1A Form

APPENDIX 6
Comparison of R&D Totals in the NSF R&D Survey and in the Economic Censuses Auxiliary Establishment Survey

This report compares the research and development data (R&D) collected in two surveys: the National Science Foundation Survey of Industrial Research and Development (NSF), and the survey in the Economic Censuses series, the Auxiliary Establishment Report (ES-9200). Both surveys ask for total expenditures on R&D in the survey year. The two years compared are 1982 and 1987.

Information gained from comparing these two datasets is limited if one is using it to verify R&D total expenditures across the surveys. By looking at the survey publications two main problems which affect the comparison become apparent: (1) the unit of analysis does not match for the two surveys and (2) the information requested varies to a degree. The comparison for most firms adds very little to verifying R&D total expenditures. However, if the goal in linking the data is to assemble a database of all establishments of firms performing R&D, for instance, than these two surveys used in conjunction provide useful information. These data are also useful in providing the location of some R&D activity. In addition for 1987 both datasets contain name and address information which helps in verifying matches across the two datasets. The present report limits the analysis to a direct comparison of the total R&D expenditures across the two surveys.

The Organizational Unit Surveyed

The first consideration when comparing these two datasets is the organizational unit surveyed. NSF surveys a firm (company or enterprise) which is defined as a "business organization consisting of one or more establishments under common ownership or control." In the Economic Censuses the unit of analysis is an auxiliary establishment of a firm. An auxiliary establishment is "an establishment primarily engaged in performing management, supervision, general administrative functions, and supporting services for other establishments of the same enterprise." An establishment such as an R&D laboratory whose principal activity is research and development is an auxiliary establishment.

The Economic Censuses unit is a subset of the NSF unit. This becomes clear when matching the data across these two surveys. Often, several observations in the Economic Censuses survey link with one firm in the NSF survey.[1] This means that

[1] We linked the two datasets by matching numeric firm identifiers or alpha codes. We only linked establishments in the NSF file that are multi-establishment firms because by definition

some firms have more than one auxiliary establishment. These auxiliary establishments may or may not perform R&D.

Table A-1 shows the total number of firms and establishments that match between the two datasets for 1982 and 1987. The ES-9200 survey contains many more observations than the NSF survey. When the datasets are matched about 55 percent of the NSF firms match to around 35 percent of the ES-9200 establishments in both years.[2] From this set, another dataset is produced that contains firms with auxiliary establishments that perform R&D. This cuts the total number of firms in the set in half to 701 and 652 firms for 1982 and 1987, respectively. A subset of the ES-9200 establishments is establishments whose principal activity is research, development, and testing or R&D laboratories (Item 7 on the survey form). For 1982 and 1987 there are 829 and 852, respectively, R&D labs.[3]

Definition of Research and Development Expenditures

The other large difference between these surveys is the instructions about what is to be included in total R&D expenditures. Both ask for "costs incurred for R&D." The NSF survey in addition to the question on the survey form also provides further explanation of what expenditures to include or exclude in an instruction manual accompanying the survey form. The Economic Censuses survey provides no additional instructions.

The lack of more defined R&D expenditures in the ES-9200 survey may lead to expenditures being included (excluded) which are included (excluded) in the NSF survey. For instance, NSF explicitly instructs respondents to exclude capital expenditures whereas no such instructions are given on the ES-9200 form. This problem would appear when a firm does all of its R&D in auxiliary establishments and the totals do not match between the two surveys.[4]

---

the auxiliary establishments in the Economic Censuses files are from multi-establishment firms.

[2] Some auxiliary establishments performing R&D may not match to an NSF firm because the firm is not selected to the NSF sample in that particular year.

[3] The Auxiliary Establishment Report reports 1086 and 1167 establishments in 1982 and 1987, respectively, that report research, development and testing as their principal activity .

[4] The totals are more likely to match when the same person in a firm fills out both forms.

Comparing Total R&D Expenditures

First I compare the universe of both datasets. As shown in table A-1, $59 and $96 million dollars were spent on R&D in the U.S. in 1982 and 1987, respectively.[5] R&D taking place in auxiliary establishments represents 27 and 21 percent of this total, respectively. About 17 and 13 percent of the total takes place in R&D laboratories. The ES-9200 represents a good portion of total U.S. R&D.

Those NSF firms that matched to firms in the ES-9200 represent 34 and 29 percent of the total R&D in 1982 and 1987, respectively. To get a better idea of how these two sets compare at the firm level I calculate mean R&D expenditures for the NSF firms and the ES-9200 data aggregated to the firm level.[6] For the average firm the proportion represented by the ES-9200 firm remains the same as the figures derived from the totals.

For the matched observations across the NSF and ES-9200 surveys, the R&D lab total decreases from 829 to 743 in 1982 and from 852 to 706 in 1987. As explained in footnote 2 this may be due to firms with labs not being included in the NSF sample. The lab total is reduced again when I limit the set to NSF firms with auxiliary establishments that perform R&D. This indicates that some R&D labs report no R&D expenditures. I have no explanation for this situation.

The linking of these two datasets provides limited information that would improve the data in the NSF survey. It does provide information about where some firms perform their R&D activity. Others may find additional applications. For this report I find that the match fails to add significant insight into total R&D expenditures.

---

[5] NSF(1989,1984).

[6] To compare the R&D total expenditures all establishment data from the ES-9200 is summed to a single record for each firm.

TABLE A-1
Comparison of R&D Data from the NSF Survey and the Economic Censuses' Auxiliary Establishment Survey.

	1982		1987	
	NSF[1]	ES-9200	NSF[1]	ES-9200
All Observations:				
Total Observations	2387	35986	2594	38236
Number of R&D Labs	NA	829	NA	852
Total R&D[2]	$58,960[3]	$16,132	$96,305[3]	$19,759
Percent of NSF R&D		27%		21%
Total R&D for Labs[2]		$10,256		$12,532
Percent of NSF R&D		17%		13%
Matched Observations Across NSF and ES-9200 Surveys:				
Number that match	1313	13674	1443	12839
Percent of Total Observations	55%	38%	56%	34%
Number of R&D Labs		743		706
NSF Firms with Auxiliary Establishments that perform R&D:				
Number	701	11766	652	10068
Number of R&D Labs		673		544
Total R&D[2]	$46,356	$15,737	$67,276	$19,309
Percent of NSF R&D		34%		29%
Percent of Total Observations	79%	98%	70%	98%
Mean R&D (per firm)	$66	$22	$103	$30
Percent of NSF R&D		34%		29%

1 Only multi-establishment firms are included here.
2 All dollar figures are in millions.

3 This total includes all R&D performed at single and multi-establishment firms. This number comes from the NSF R&D publication.