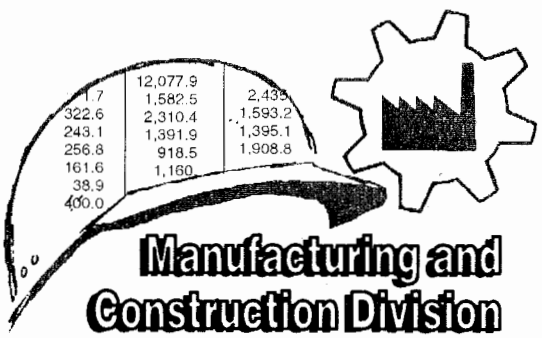
Working Papers

MANUFACTURING AND CONSTRUCTION DIVISION
REPORT SERIES
MCD Working Paper Number:
Census/MCD/WP-95/01

DESIGN OF THE SURVEY OF INDUSTRIAL RESEARCH AND DEVELOPMENT: A HISTORICAL PERSPECTIVE

by
Gary L. Kusch
William Ricciardi
Manufacturing and Construction Division
Bureau of the Census

Industrial and Construction Statistics



BUREAU OF THE CENSUS
MANUFACTURING AND CONSTRUCTION DIVISION
REPORT SERIES
MCD Working Paper Number:
Census/MCD/WP-95/01

DESIGN OF THE SURVEY OF INDUSTRIAL RESEARCH AND DEVELOPMENT: A HISTORICAL PERSPECTIVE

by
Gary L. Kusch
William Ricciardi
Manufacturing and Construction Division
Bureau of the Census

June 1995

Introduction

The Survey of Industrial Research and Development (R&D) is a company-based survey conducted by the Census Bureau for the National Science Foundation (NSF). The survey collects detailed information on R&D performance and expenditures for both manufacturing and nonmanufacturing companies that operate for profit.

Between 1953 and 1956, two surveys measuring R&D performance were conducted by the Bureau of Labor Statistics (BLS), and since 1957, the Census Bureau has been conducting the R&D survey. Over this span, the survey has provided a continuum of data on R&D expenditures by major industry group by source of funds. Questions on additional characteristics, such as product field, energy, and pollution abatement, have been added over the years as our focus on R&D activities has expanded. Response to four questions (total net sales and total employment for the company; the amount of Federal and total funds the company spent on R&D) are mandatory under Title 13, U.S. Code, sections 131, 182, 224, and 225, however, the remaining questions on the survey are voluntary.

In order to evaluate and improve the quality of the R&D survey and to introduce design enhancements, the NSF funded an improvement initiative project to be conducted by the Census Bureau during FY 1993-1995. This project consists of five tasks, one of which was designed to evaluate and improve the survey's sample design and methodology. One part of this task was to prepare a document that chronicled the survey design activities of the R&D survey during the time it has been conducted by the Bureau. This report is the result of that effort.

The objective of this historical documentation is to examine those issues mostly affecting the survey design. This goal required us to consider many aspects of the survey, but the document is not intended to be a comprehensive survey history. Many important issues, such as response rates, use of multiple forms, mandatory-voluntary arguments, and areas of subject-matter expertise are not addressed at all or are addressed only in brief. A historical treatment of such issues is certainly worthwhile but falls outside the scope of this report.

The primary sources were past R&D publications, recent Office of Management and Budget (OMB) packages, and Census Bureau specifications that describe in detail many of the methodologies employed in the R&D survey used for this report. The R&D publications described many activities, such as company classification and sampling methods, only in the most general of terms. Census Bureau documentation prior to 1981 is no longer extant. Thus, in many instances we cannot provide exact details of the survey design. Throughout the report we have attempted to point out those areas where our information is incomplete.

Presentation Outline

A total of eight survey periods are considered: 1957-1961, 1962-1966, 1967-1970, 1971-1975, 1976-1980, 1981-1986, 1987-1991, and 1992-1994. These periods, in general, correspond to the introduction of new sample panels for the survey.

Each survey period is discussed in relation to the following five categories: scope, frame source, classification, sample design, and unique problems and issues. We present these topics in this order, not to suggest relative importance, but to parallel issues considered in arriving at an overall survey design for the R&D survey.

In general, for any survey design, the first step is to define the objectives and scope of the survey. Once the scope has been determined, the sources available for constructing a sampling frame and the information on those sources are a logical second step to consider. The frame can originate from a single source, or, as with the R&D survey, the frame can be constructed from many different sources. Classification issues for the R&D survey are extremely important since it is a company-based survey and many different ways of assigning company Standard Industrial Classification (SIC) codes can be considered. In fact, the method of classification used in the R&D survey has changed over the years. No matter what method of classification is employed, the assigned SIC code is intended to represent the primary activity of the company, regardless of how primary activity is defined. The main purpose of the classification code, however, is to group companies together for tabulation purposes.

Although sample design decisions for the R&D survey usually are made early in the planning process, we address sample design after classification because some sample design issues (for example, stratification, levels of control, sample allocation, etc.) are dependent upon the results of classification. Finally, we included topic five because unexpected problems or issues which have potential sample design implications can arise at any time during the survey planning process or during the conducting of the survey. Sometimes it is possible to resolve these issues at the time and these resolutions can be reflected in the current sample design. At other times, special studies or evaluations may be required, and results from these studies may not be implemented until the next survey design takes place.

Attachment A to this document is a historical chart highlighting major changes relative to the five factors for each of the eight survey periods.

1957-1961

Scope: The Census Bureau conducted the R&D survey for the first time in 1957 and a survey panel was selected which remained in effect, with modifications, through the 1961 survey year. During these years, R&D activities in the industrial sector were impacted significantly by the increased importance of science and technology to national security considerations.

The survey in 1957 covered all manufacturing industries and selected nonmanufacturing industries (communications, radio and television broadcasting, and crude petroleum and extraction industries). Organizations known as Federal contract research centers, later known as Federally funded research and development centers (these were organizations engaged in R&D projects for the Federal Government through a contract), were excluded from the 1957 survey, but they were included in all subsequent years. Trade associations, being nonprofit, were excluded for all years even though they did perform a certain amount of R&D. Agricultural cooperatives also were excluded. In 1958, for the first time, a representative sample of companies covering almost all nonmanufacturing industries was introduced. Those excluded from the nonmanufacturing sample were the railroad industry (due to its presumed lack of R&D) and nonmanufacturing companies with less than 50 employees. The majority of the changes to scope during this period took place within the first 1 or 2 years the Census Bureau conducted the survey.

Throughout this period two basic forms were used. A detailed form (RD-1) was used for the larger R&D performing companies and an abbreviated form (RD-2) was used for the smaller companies. Large companies not responding were mailed Census Bureau mandatory form MA-121 to collect net sales and receipts, number of employees, and cost or receipts for R&D by Federal funds and company funds.

Frame Sources: In this first survey period, a new sample (panel) of roughly 6,800 companies was drawn for the R&D survey for use beginning with the 1957 survey year. The frame for this sampling operation was constructed from several sources, the primary ones being the Census Bureau's Annual Survey of Manufactures (ASM) and files from the Bureau of Old-Age and Survivors Insurance (BOASI). These remained the basic frame sources for the next R&D sample panel, survey year 1962.

The sample panels were supplemented in nonsample years in a variety of ways. Lists obtained each year from the Department of Defense (DOD) containing the largest R&D contractors were matched against the panel, and companies not already in the panel were added as certainty cases regardless of industry classification or employment size. In 1958, the panel was augmented by an operation in which any multiunit company with 1,000 or more employees and not in the current panel was added. The panels were further supplemented from lists of newly-formed businesses, or births.

Classification: An integral part of the survey design process is to assign each company in the frame a single classification. The classification is intended to reflect the major activity of the company and serves as the basis for tabulating survey data by industry group.

Throughout the history of the R&D survey, the term "major activity" has represented different things. When the Census Bureau first conducted the survey, the major activity of a company was defined in terms of value added from the ASM. In later years, to the best of our knowledge, major activity was based upon product class shipments. For the most recent survey designs, the major activity of a company was determined using its total company payroll. In none of these situations was it assumed that the assigned code indicated where company R&D activity was concentrated. Company codes based on R&D information could not be assigned since this information is never available for the entire frame. Therefore, it was necessary to classify a company based on available information. Even if R&D data were available, it is not clear that this is the intent of the company classification, especially since Item 8 attempts to identify R&D by end product. The intent of the code may be more general-purpose so that R&D survey data can be compared to data from other company surveys. Such comparisons are limited if the methods of classification differ substantially in concept.

Industry groupings, which were based on the set of classification codes, defined the primary sampling strata used in the sample design. In this period, the number of industry groupings varied. There were 18 groupings in 1957 (16 manufacturing and 2 nonmanufacturing), but this was expanded to 22 in 1958 (21 manufacturing and 1 nonmanufacturing). Thereafter, the groupings remained unchanged except for 1960 when stone, glass, and clay was included in the "other manufacturing" category. In 1961, stone, glass, and clay was treated as a separate industry grouping again.

Classifying single-unit companies was relatively easy since they were engaged in one activity. Multiunit companies, however, were often engaged in activity in more than one industrial area, so it was necessary to develop specific rules for determining their classification for this survey. Documentation for that time does not provide definitions of these rules.

The classification code, once assigned, generally was subject to change only when newer information from the 1958 Economic Censuses became available. Documentation is not detailed, but it is apparent that more frequent changes to codes were allowed. In 1957, a company was classified in manufacturing, mining, or other industrial area based on employment taken from the ASM. Within manufacturing, companies were coded on the basis of value added.

Nonmanufacturing companies were coded on the basis of BOASI industry codes. In either 1958 or 1959, a slight change was introduced by which a company in manufacturing was classified, as best we can infer from the documentation, based on its product class shipments. In both these years, changes also were made for a few large companies involved in mergers and acquisitions. A year later, in 1960, more significant changes in classification were introduced with the release of the 1958 Census. We assume that the same rules applied in the original coding were followed, but we cannot state this with certainty due to incomplete documentation.

Changes in company classification led to partial revisions to the R&D data series for the years 1958-1960. The 1958 revision reflected code changes only for companies with 1,000 or more employees. Revisions to the 1959-1960 data were for certain industry and size groups only. By the time of the 1961 survey, all coding changes based on the 1958 Census had been effected. Thus, no subsequent revision to this data year was necessary.

The 1961 survey year also marked the last year for the sample panel. A new sample panel was selected for use beginning with survey year 1962, the first year in the survey period to be discussed next. Company codes for the new panel were still based on the 1958 Census, and so companies overlapping both panels had no change in code. When results of the 1963 Economic Censuses were available, these companies (as well as the nonoverlap companies of the new panel) were recoded. The R&D data series for 1959-1962 were then revised based on code changes for overlap cases. We describe the methodology employed in this revision in the "Classification " section of the 1962-1966 survey period. We mention it here because it resulted in a revision to the data series for this period.

One other point should be kept in mind regarding coding changes in this period. A potential source for some of these changes was the revision to the classification system used by the Census Bureau. Until 1960, a company was assigned an industry code based on the 1947-49 editions of the SIC manual. In 1960, the 1957 SIC manual was used to assign industry codes.

One other source of revision, independent of those arising from coding changes, has occurred on a yearly basis since the time the survey has been conducted by the Census Bureau. For any given survey period, the respondent was provided with information reported in the previous survey period. This system of having previous data available on the same form that the company would use to report current data is the essence of a "shuttle" reporting system. The respondent was asked to make changes to this prior data, which were preentered on the form by the Census Bureau, to make them comparable to the data reported in the current year. Such changes were made, for example, to reflect changes in reporting concepts or changes in company structure, such as mergers or acquisitions.

A revised prior period estimate based on these changes appeared in the publication for the current year and became part of the historical series. This operation insured substantial comparability over any 2-year period of the survey.

Sample Design: The sample design for the R&D survey during this period was based on a stratified random sampling scheme. It appears the sample, first drawn in 1957, was augmented in later years with companies added as a result of additional sampling operations. These operations may, in fact, have been no more than panel expansions (perhaps due to an expanded sampling frame, or to the desire to completely canvass certain strata that were previously sampled) on the one hand or panel contractions (subsampling) on the other. In any event, they clearly were not complete resamplings of the survey.

During the primary sampling operation in 1957, the universe was stratified by industry classification and employment size. The sampling unit was the company, defined as all establishments under common ownership or control. Establishments of a company were assigned a probability of selection. Thus, single-unit companies were selected with probabilities of selection equal to that of the corresponding establishment. For multiunit companies, the probability of selection was set to the sum of the probabilities of their respective establishments. Once the final probability of selection was determined for each multiunit company, they were selected by a random process in accordance with their assigned final probabilities. The actual sample selection process was not described clearly in the documentation, but we assume it followed either a systematic or independent random approach based on assigned probabilities. We found no reference in the documentation about what level of reliability was sought. We believe that there was some desire to relate the size of the estimates for the industry groupings to their respective relative standard errors, (for example, as the estimates increase the relative standard errors decrease). Certainty cases in the sample accounted for almost 90 percent of the total R&D performance funds, so this probably ensured that reliable estimates were obtained for most publication cells.

Sampling fractions for some strata varied over the years, so apparently, modifications to the existing panel were being made. For example, in 1957, a sampling fraction less than 1.0 was employed for manufacturing companies with 1,000-2,499 employees.

But in 1958, the previous panel was expanded to include all companies in this stratum with weights of 1.0 in order to provide greater comparability with the Survey of Scientific and Technical Personnel, a BLS survey. Companies reporting no R&D in 1957 were retained in the 1958 survey if they were in a certainty stratum in 1957; but if they were in a noncertainty stratum in 1957, they were sampled at a reduced rate for 1958. These situations apparently represented specific panel enhancements or subsampling operations and not a complete resampling of the survey. In some instances, they probably were made in an effort to improve the relationship between the estimate and its associated standard error.

The manufacturing companies (1957-1961) were drawn primarily from the 1956 ASM. The ASM was not a manufacturing universe; it was a sample panel representing the manufacturing universe. What is not evident from the discussion above regarding probability assignment was whether the probabilities as described represented a second-stage assignment only or some composite probability that took into account the first-stage ASM sampling. We must assume that the status of the ASM as a sample was accounted for. The nonmanufacturing companies (1957-1961) were drawn from the 1956 BOASI records. Those nonmanufacturing companies that reported no R&D when they were originally sampled were not included in subsequent surveys.

Unique Problems/Issues: The BLS conducted the R&D survey from 1953 to 1956 prior to the Census Bureau takeover in 1957. In order to evaluate the comparability of its estimates with the BLS estimates, the Census Bureau collected 1956 data in the 1957 survey. The BLS estimates were about 4 percent higher than the Census Bureau estimates for 1956 and even higher for some industry groupings. Reasons for these differences included: (1) some companies used different methods for computing figures for the two surveys; (2) the Census Bureau used a 1956 list for the 1957 sample, whereas the BLS used a 1951 list for the 1956 sample; (3) industry codes assigned to a few important companies differed between the two surveys; and (4) the phrasing of questions asked of small companies differed in the two surveys. These differences made it unlikely that consistent trend data between the two series would have resulted for individual industries for 1953-57. As a result, the NSF adjusted the estimates for 1953-55 using a relationship between 1956 BLS and Census Bureau data. These adjusted estimates, together with data for 1956 and the 1957 survey, provided a basis for developing a consistent 1953-57 time series.

Another issue surfaced in 1958 when the relative share of R&D performance that was financed by the Federal Government may have been understated in certain industries. Discussions held with respondent companies indicated this may have occurred because some company-related R&D was indirectly financed through overhead payments under Federal contracts. It often was difficult for companies to separate the portion of funds for company-related R&D from that which was Federally financed through overhead charges.

Problems such as this arose from a lack of accounting and estimating procedures that allowed a company to provide the information requested. In some companies the importance of R&D to the company encouraged the development of improved record-keeping procedures for R&D activities.

A second issue addressed in 1958 dealt with the difficulty in identifying R&D activities in small firms. A supplemental sample of roughly 4,000 small firms was drawn to obtain estimates for two small size classes: 0-7 employees and 8-99 employees. A sampling variability of approximately 10 percent was targeted. These cases were treated as a separate sample of manufacturing companies, which permitted separate estimates to be formed for the original sample and for the supplemental sample. The desired reliability figures for these small company classes were not obtained. The conclusion drawn from this experience was that there were still considerable concept (response) errors to be isolated. Also, information obtained from the reporting of small companies confirmed earlier suggestions that attempts to measure the year-to-year change in small company classes would be difficult. However, the test did provide level estimates for these small-company classes.

1962-1966

Scope: As far as we can determine, no changes in scope occurred in this survey period. The survey still included all manufacturing industries and nearly all nonmanufacturing industries. The railroad industry continued to be excluded due to its lack of R&D, and nonmanufacturing companies with less than 50 employees continued to be excluded as well. Industry-oriented organizations, such as trade associations and agricultural cooperatives, still were not covered.

The basic forms, RD-1 for large R&D performing companies and RD-2 for small R&D performing companies, continued to be used. In order to collect basic information, mandatory form MA-121 continued to be mailed to large companies that did not respond initially. New in 1962, was the introduction of an attachment form to the basic RD-1 form. This attachment, which in 1967 would be known as the RD-11 form, was sent to companies reporting more than $1 million in Federal research and development. These companies were asked to provide separate figures for the three categories of Federal agencies: (1) DOD; (2) National Aeronautics and Space Administration (NASA); and (3) all other Federal agencies.

Frame Sources: For survey year 1962, a new panel of approximately 7,000 companies was selected. The frame from which the panel was selected was constructed from the same basic sources as in the previous panel selection. The primary frame sources were the Census Bureau's ASM files and the BOASI files. Panels were still supplemented in nonsample years by a list from DOD of the largest R&D contractors.

Classification: The assignment of a company classification for the 1962 sampling operation was based upon company information obtained from the 1958 Economic Censuses. As with the earlier panel, we are unable to specify what specific rules were applied to classify multiunit companies. We believe, however, that they were the same rules used in the original coding for the first panel and in the recoding of that panel in 1960 when the 1958 Census first became available.

The new sample panel underwent a similar recoding in 1965 when the 1963 Economic Censuses data were released. Again, we believe that the same rules of classification were followed. We also assume that mergers and acquisitions by large companies were monitored in each of these survey years as in the earlier period, and that company classifications may have changed as a result. The number of industry groupings remained at 22 for the duration of this period. The breakout of the groupings (21 manufacturing and 1 nonmanufacturing) was consistent over all 5 years.

The use of the 1958 Census in the original coding for the new survey meant that companies common to both the old and new panels had the same classification each time. As a result of the recoding in 1965, however, many of these companies received a different code. This indicates that in one of the years between 1958 (the basis of the original code) and 1963 (the basis of the new code) the company's primary activity changed. The R&D data series for 1959-1962 were revised to account for the changes in overlap cases. Since for any given company it was not possible to know in which year the switch occurred, a methodology was adopted which made no assumption about the year of the switch. Instead, for each year, the company's data were allocated between the two codes in changing proportions. Thus, in 1958, all of the data remained in the original code; in 1959, 80 percent of the data were allocated to the original code and 20 percent to the new code; in 1960, 60 percent of the data were allocated to the original code and 40 percent to the new; in 1961, 40 percent were allocated to the old code and 60 percent to the new; in 1962, 20 percent were allocated to the old code and 80 percent to the new; and finally, in 1963, all the company's data were allocated to the new code.

The revised series first appeared in the 1965 publication of historical tables. The 1964 and 1965 data years, originally tabbed based on the 1958 codes assigned to these companies, were retabbed to reflect the updated codes. The 1966 survey data were collected based on the new codes, so no retabbing was necessary for that year. Each of these latter 3 years would be revised (or in some cases re-revised) when 1967 Census data were made available in 1971. The discussion on these revisions is deferred until the 1971-1975 period is addressed later.

As discussed before, independent of revisions due to reclassifications, yearly revisions to prior period estimates resulted from changes being made by respondents to their prior period data. These revisions ensured substantial comparability over any 2-year period.

Sample Design: No major sample design changes occurred for this sampling operation. The design remained a stratified sample with stratification by industry grouping and employment size. Probability assignments for multiunit companies again were based upon probabilities of selection for the individual establishments of the company. A random selection process, based on the assigned probabilities, was performed for each of the strata. No mention was made in documentation that particular reliability constraints were being met. We believe that a relationship between size of the estimate and its relative standard error was sought. Certainty companies (about 1,800 in number) accounted for about 95 percent of total R&D performance funds.

The ASM served as the primary frame for the manufacturing universe. As before, we assume the status of the ASM as a sample, and not a universe, was accounted for in the ultimate probability assignment for a company or in the weighting of company data when estimates were formed. The BOASI records remained the prime sampling source for the nonmanufacturing records.

There are no indications that the sample panel was augmented in later years as occurred for the earlier panel. However, companies that reported no R&D activity in the first mailing were mailed in subsequent years on a subsample basis. Thus, the number of mailed cases declined after the first year.

Unique Problems/Issues: In the 1962 survey year, the results obtained for two questionnaire items were not published because response to the item was either lacking or suspect. A company's inability to supply information from their existing records was believed to be the primary cause of this situation. Items that experienced response difficulty were "Forward Budgeting of Company Funds for Research and Development Performance Within the Company," and "Capital Expenditures for Research and Development." For the "Forward Budgeting" item, fewer than 20 percent of the large companies responded. After the first year, the data became more sparse, with fewer than 10 percent of the companies reporting.

Capital expenditures for research and development for most companies were included in company accounts that covered more than research and development, thus making it harder to report the requested item. Only about 20 to 30 percent of research and development costs were reported by separately organized research and development laboratories. As a result of these findings, these items were removed from the questionnaire pending further investigation of methodology and techniques for obtaining more accurate estimates for them.

The problem of identifying R&D activity in companies with fewer than 100 employees was confirmed again when the new sample was drawn in 1962. Estimates prepared from the current sample (1962) and from previous samples for this small-company class were shown to vary significantly. For example, the 1957 sample estimated that there were 6,800 small companies conducting R&D, whereas the 1958 supplemental sample and the 1962 sample gave estimates of roughly 12,500 and 12,000 companies conducting R&D, respectively. It also should be noted, however, that the sample design for the R&D survey was not meant to provide good estimates of company counts, but rather the amount of R&D being conducted. This was probably a major reason why the estimates for small company counts varied so much.

1967-1970

Scope: The scope of the R&D survey for this period did not change. The survey continued to cover for-profit companies of all manufacturing industries and nonmanufacturing industries believed to conduct or finance research and development. Nonmanufacturing companies below 50 employees still were excluded.

A basic change in the treatment of the small R&D-performing companies was initiated in survey year 1969. These firms, heretofore mailed short form RD-2 annually, were not mailed in 1969 or 1970. Data were imputed for these companies in each of these years. In subsequent survey periods, companies initially identified as small and eligible for the RD-2 form, were mailed only the first year of the survey panel. Unless they were redefined as being large that first year, these small companies were not mailed in subsequent years, but were imputed instead. The imputation was done by applying industry-level year-to-year ratios to a company's prior period data. In the first year of imputation, these ratios were applied to actual reported data, whereas in subsequent years, these ratios were applied to imputed data. Mandatory form MA-121 continued to be mailed each year to large companies not responding initially.

Frame Sources: For survey year 1967, a new sample panel of approximately 8,000 companies was drawn. New sources were used in constructing a frame for this sampling operation. The 1963 Census Enterprise Statistics file was the source for identifying multiunit manufacturing companies. This file included company-level information obtained from the 1963 Economic Censuses. Single-unit manufacturers were identified from the 1963 Census of Manufactures file. The nonmanufacturing universe was developed from a 1966 file of Social Security Administration (SSA) records. The selected panel continued to be supplemented in nonsample years from DOD lists of R&D contractors and also from a list of R&D contractors provided by NASA.

Classification: The rules for determining multiunit company classifications were, to the best of our knowledge, unchanged from previous survey periods, but as before, no detailed information is available. Classifications were based on the 1967 SIC manual and not the 1957 manual that had been used previously, but this change likely had little effect on coding. No other relevant changes took place in this period. Since data from the 1967 Censuses were not available during this span, no recoding operation similar to the one described in the 1962-1966 period was required. Census Bureau data (1967) were available by the next survey period (1971-1975), and a recoding at that time did result in revisions to the 1967-1970 R&D series, as well as to the years 1964, 1965, and 1966 of the previous period. During each survey year, revisions to prior period estimates based on changes made by respondents to their prior period data continued to be made.

The number of industry groupings increased to 23 for survey year 1967 (the radio and television receiving equipment industry was split out from other electrical) and remained at that total for each of the years in this period. As before, all but one of these groupings were manufacturing.

Sample Design: The basic sample design remained unchanged for this sampling operation. Stratification by industry grouping and employment size continued, and probabilities of selection for multiunit companies still were based upon the individual probabilities assigned to establishments of the company. A random selection process, based on the assigned probabilities, was performed within each stratum. Again, it was not stated whether particular relative standard error constraints were established. We believe, as before, that the ultimate goal was to ensure a relationship between the size of the sample estimate and the relative standard error of the estimate. Reasonable estimates at most levels likely were assured because of the continued high coverage (almost 95 percent of total R&D performance funds) of the certainty companies.

Perhaps the most significant change in the operation was the fact that the sample frame for manufacturing, as defined by the Enterprise Statistics files and the census of manufacturers file, was a complete representation of the universe. In previous sample years, the manufacturing frame was constructed using the ASM. The ASM was itself a sample panel. This change removed any complications arising from the fact that the ASM had to be treated as a first stage of sampling in the design of the R&D survey. The use of updated 1966 SSA records as the source for the nonmanufacturing sample frame portion also represented a considerable improvement. The BOASI records used in previous operations dated to 1956.

Unique Problems/Issues: The data collected from the new panel in survey year 1967 suggested, as in 1958 and 1962, that there was a problem in identifying R&D in small companies (for these purposes, companies with employment less than 100). The estimated number of companies in this size class varied over these 3 years: about 12,500 in 1958 from the supplemental sample, about 12,000 in 1962, and about 10,000 in 1967. The 1957 sample estimate of about 6,800 emphasized this variability. As was mentioned with regard to the previous survey period, a large part of this variation must be attributed to the fact that the sample designs were not optimal for estimating company counts. The primary objective of the designs was to estimate the amount of R&D. While small companies accounted for almost 90 percent of the total number of companies, they accounted for less than 0.5 percent of the total R&D funds for all companies. The contrast between these 2 percentages supports the contention that a design to estimate total R&D funds would not provide good estimates for company counts.

1971-1975

Scope: During this period, no major changes in scope occurred. Nonmanufacturing companies with employment below 50 continued to be excluded.

No change in the use of forms RD-1 and RD-2 occurred in these years. The large R&D performing companies continued to receive form RD-1 each year. Companies initially designated to receive the short form RD-2 were mailed only the first year of the panel unless they were converted to the long form based on the amount of R&D reported that first year. In subsequent survey years, small companies not mailed were imputed as described earlier. New to this survey period was the addition of the RD-3 form. This form, first used in 1974, tested the feasibility of collecting scientific and technical information expenditures data. The item was reworded for the 1975 survey due to companies having difficulties providing the requested data. Similar reporting problems continued in 1975, and the form was dropped prior to the 1976 survey. Also, last appearing in the 1975 survey was the RD-11 form. It was first used in 1962 for companies reporting more than $1 million in Federal research and development.

Frame Sources: A representative sample of approximately 8,000 manufacturing and nonmanufacturing companies was selected for use in the 1971 survey year. This new sample was based on frame sources which, for the most part, were updated versions of the sources used in 1967. Thus, multiunit manufacturing companies were identified from the 1967 Enterprise Statistics file and single-unit companies from the 1967 Census of Manufactures file. In a slight departure from 1967 sampling, the Enterprise file also was used as the frame source for selected nonmanufacturing industries (SICs 7391-92, 7397, 8911). The remaining in-scope nonmanufacturing industries again were identified from SSA records.

Lists of large R&D contractors continued to be obtained each year from DOD and NASA, and companies on these lists not on the current R&D panel were added as certainties.

Classification: Rules for classifying multiunit companies were not in the documentation available to us. We assume that no changes occurred from prior years. The number of industry groupings published remained at 23 for 1971 and 1972, but was increased to 25 for the years 1973-1975. Electrical components and other transportation equipment were treated separately. As before, all but one industry grouping was manufacturing.

Classifications for overlap companies between the new and old panels were sometimes different since the codes were based on company information from two different points in time (1967 for the new panel and 1963 for the old). The same was true for overlaps between the new panel and the 1962 panel. One can assume that these companies changed their primary activity during one of the intervening years.

The R&D data series for years 1964-1966 and 1967-1970 were revised due to the new classification. The intervening years were revised using the same methodology described earlier (see Classification section for the period 1962-1966). However, since there were only 4 years involved in this revision, the allocation factors were based on multiples of 25 percent, and not 20 percent as before. Thus, for example, for 1963, all of an overlapping company's data were tabbed in the old code; in 1964, 75 percent were allocated to the old code and 25 percent to the new code, etc. Note that this represented a second revision to the 1964 and 1965 data years, and a first revision to the 1966 data year. The subsequent years (1967-1970) were revised by simply retabbing these years using the updated codes for the overlap companies. This marked a first revision for these years. They were subject to a second revision once a new panel was drawn for survey year 1976. Details of this revision are deferred until survey period (1976-1980) is discussed.

As usual, the yearly estimates provided by the new panel (1971-1975) were revised each year based on changes made by respondents to their prior period data. Additional revisions for most of these years (1972-1974) were made later and will be discussed in the next survey period.

Sample Design: There were no changes in sample design from the previous survey period. A stratified random sampling scheme was utilized as before to select the 1971 sample panel.

Unique Problems/Issues: A "response analysis" study was conducted jointly by the Census Bureau and the NSF in 1975. The purpose of this study was to discuss each questionnaire item in detail with respondents to determine the sources used by companies to provide data, to examine their methods of estimation, and to identify problems encountered by respondents. Over 100 interviews were conducted. Firms classified in a wide variety of industries were selected, but there was an emphasis on the larger R&D performers in the more R&D-intensive industries. Questionnaire items relating to scientific and technical information expenditures and company research and development as indirect costs of Federal contracts were found to be difficult to answer and were removed from the survey. Other items were simplified or deleted due to the lack of reliable data, and additional instructions were written for some items. The Census Bureau and the NSF worked with the respondents by offering suggestions, providing feedback, and revising the questionnaire.

1976-1980

Scope: In this survey period, no change in the scope of the R&D survey occurred.

A significant change in the survey operation was incorporated during this period, however, to further address the issue of reporting burden. Heretofore, companies in the survey panel identified as large in terms of R&D were mailed a detailed questionnaire (RD-1) each year. Beginning in 1978, the large companies began alternating between a long-form version (odd years) and a short-form version (even years) of the RD-1 questionnaire. Small companies continued to be mailed only the first survey year (1976) using short form RD-2 (later to become form RD-1A). Unless they were converted to large-company status as a result, they were imputed in subsequent survey years (as discussed earlier).

Frame Sources: A significant change in frame sources occurred in the sampling operation for the 1976 survey panel. For the first time, the Census Bureau's Standard Statistical Establishment List (SSEL) was used. The SSEL is an annually-updated master file comprised of all nonfarm economic entities. The 1974 version of the SSEL was the prime frame source for all manufacturing industries and for selected nonmanufacturing industries (SIC 49, 7391-92, 7399, 8911) for single-unit companies. Records from the SSA were used to identify single-unit companies in the remaining in-scope nonmanufacturing industries. Multiunit companies were identified from the 1972 Enterprise Statistics file. The total number of multiunit and single-unit companies selected for the 1976 sample was approximately 11,500.

Classification: Based on available documentation, there is still no mention of rules for determining the classification of multiunit companies. We assume the rules have not changed, but we are not assured that this is so. This is because the SSEL, compared to the prime sources used in many of the previous sampling operations, has limited data information available for classification purposes, and it might not have been possible to apply the previous procedures exactly. The number of industry groupings was increased to 26 in 1978. Other nonelectrical machinery was treated as a separate manufacturing grouping. The 1972 SIC manual had replaced the 1967 manual by this time and was the basis for assigning codes to the establishments in the source files.

Once again, with the introduction of a new panel, companies which overlapped the old and new panels often had different classifications since company information from two different sources and points in time were utilized in the respective codings. The usual assumption was made that the primary company activity changed between 1967 and 1974. The same methodology described for the periods 1962-1966 and 1971-1975 was applied to revise the R&D data series for these intervening years. This time, however, because of the number of years involved, the allocation factors applied to overlapping companies were multiples of 14.3 percent.

Further revisions to the R&D data series were introduced during the first year of the new panel (1976). In the 1976 survey year, prior-period data were collected for the entire panel. Thus, a new panel estimate for 1975 was generated and compared to the old panel estimate. This new estimate was considered superior since it reflected updated coding and was based on a new panel that had not undergone the deterioration of the old panel. As a result of this comparison, panel estimates for the middle 3 years (1972-1974) of the old panel were adjusted for each industry grouping. The first year (1971) did not require adjustment since it was the first year of that panel. The adjustments for each of the other years were increasing proportions of the total differences observed between the two 1975 estimates. Thus, 25 percent of this difference was the adjustment to the 1972 estimate, 50 percent of this difference was the adjustment to the 1973 estimate, and 75 percent of this difference was the adjustment to the 1974 estimate. These revisions were the first of the so-called "wedging" operations performed on R&D data.

Wedged revisions differed in important ways, both operationally and conceptually, from the major revision procedures described earlier in this report. In earlier years, revisions were made because companies common to successive panels were often coded differently. Data file records for these companies were adjusted to account for this, and the data file retabulated to generate revised estimates.

Thus, the estimates remained derivable from the R&D microdata. In 1976, wedging involved changes to estimates only. No "correction" was made to individual microdata or to any data file. Thus, the revised estimates could not be generated from the R&D microdata.

Wedging differed conceptually in the sense that it accounted for coding changes that theoretically occurred for the entire universe but were not allowed because of the freezing of company codes. The earlier procedures accounted for coding changes as well, but only for that portion of the universe represented by the overlap companies. In this respect, wedging was a preferred methodology.

For the reason stated above, wedging became the primary method of revising the R&D data series. With each new sample selection, both current and prior estimates were obtained in the first survey year of the new panel. The series obtained using the old panel were revised by wedging to the new prior-period estimates. Thus, the series from 1976-1980 were revised based on 1980 estimates obtained from the new 1981 survey panel. The wedging algorithm was modified in later applications to preserve the trend pattern of the original series to the degree possible. In 1992, with the introduction of annual sampling, wedging was discontinued because company codes were updated each year and because one-year panels suffered no coverage loss.

One other note should be made about wedging. It was not the "original" series for a given panel that were being revised. Recall that each year of a survey, each respondent was allowed to change their reporting for the prior year so that it was comparable to their reporting in the current year. Revised estimates for the prior year became part of the historical series. It was these estimates that were revised during wedging.

Sample Design: There were no major changes in the sample design from the previous survey period. A stratified random sampling scheme still was utilized to select the 1976 sample panel. The strata, 26 industry groupings by 4 total employment (TE) size classes, remained the same. Companies with 1,000 employees or more were included in the sample with certainty. Different sampling fractions for individual strata were derived based on new information with the intent of improving the relationship between the size of the estimates and their estimated relative standard errors.

Unique Problems/Issues: No unusual problems or issues arose during this period.

1981-1986

Scope: No significant changes in scope for the R&D survey occurred in this survey period. Census Bureau single-unit companies below 5 employees were part of what was called the administrative record universe. Such companies are not mailed in the census but rather are imputed from administrative records to keep response burden low for this segment of the population not expected to contribute much to R&D expenditures.

Alternating use of a long form RD-1 and a short form RD-1 continued for the large R&D performers in odd and even numbered years respectively. In 1984, the short-form questionnaire was expanded to include items on basic and applied research and on development. Companies in the 1981 panel originally designated to receive the RD-1A form (formerly form RD-2) were mailed only the first year unless they were subsequently converted to the long form based on information they reported. In succeeding survey years, the remaining small companies were imputed (as previously mentioned).

Frame Sources: A new sample of approximately 11,500 manufacturing and nonmanufacturing companies was drawn for use in the 1981 survey. Census files continued as the primary frame sources for identifying in-scope companies. In 1981, multiunit companies were identified from the 1977 Enterprise Statistics file and single-unit companies from the 1981 SSEL file. Annual lists of R&D contractors provided by DOD and NASA continued to supplement the panel throughout the survey period.

Classification: In 1981, for the first time, we can state with some precision how classifications were derived for multiunit companies. They were based on Enterprise Industrial Category (EIC) codes assigned to companies comprising the Enterprise Statistics file. The EIC codes are closely related to SIC codes and were derived basically as follows for a given company. Payroll data for the establishments of the company were summarized to determine which of the 10 major economic sectors (agriculture, forestry, and fishing; mining; construction; manufacturing; transportation and public utilities; wholesale trade; retail trade; finance, insurance, and real estate; health services; and services except health) were predominant. Within that largest sector, each establishment of the company had its SIC-3 code mapped to its corresponding EIC code. Tabulations within the company then were made to determine which EIC within the largest sector had the most payroll. This became the company EIC code. The company EIC code determined in which industry grouping or recode the company was classified. The number of published industry categories remained at 26 for this period.

In 1981, as in 1976, survey data were collected for the prior year (1980) and an independent prior-year estimate was developed. This "revised" 1980 estimate, based on the coding of the new panel, formed the link for the second wedging operation performed on R&D data. The old panel series (1976-1980), which were based on codes that had been frozen since the panel was first selected, were revised to account for lost coverage and for coding changes that would have occurred over time. The first-year estimate (1976) was not revised since no change in codes was assumed that year and coverage was complete. A slight modification was introduced to the wedging algorithm. The new algorithm preserved, to the extent possible, the trend pattern of the original series. With the previous algorithm, the incline (decline) in the revised series increased at a constant rate each year, culminating in the last year with the full difference observed between the two independent estimates.

The estimates generated by the new panel for 1981-1986 reflected a continuation of the wedged series. Since company codes remained frozen over this span, these series were wedged after the selection of a new panel for survey year 1987 and the development of an independent 1986 estimate.

Sample Design: Major changes to the sample design occurred for the 1981 sampling. The design remained stratified, but for the first time, probabilities of selection for noncertainty companies in the universe were made proportionate to a company measure of size. This is referred to as probability proportionate to size (pps) sampling. The measure of size was the reported R&D expenditures value for those companies in the frame which had reported a value in the previous panel year (1980). For all other companies, an estimate of R&D expenditures was made. R&D expenditures were estimated from total employment based upon a relationship between these two variables developed from the most recent survey data. These relationships were derived for each of the 26 industry groupings. The industry groupings formed the strata, but this time (except for the certainty stratum described below) no sub-stratification by TE size class was done. An administrative decision to include all in-scope companies with employment of 500 or more in the sample, regardless of their measure of size, was made. These are referred to as "predetermined" certainties.

With a measure of R&D (either reported or imputed) present for each company, relative standard error constraints were established on estimated totals of this measure for each industry grouping. This was done consistent with the overall sample size constraint that was imposed. The primary importance in applying these constraints was that they determined the actual probability for each company and, therefore, the ultimate allocation of the sample across industry groupings.

Since the actual distribution of "true" R&D values across the universe was different from the surrogate or imputed values, it introduced a limitation in the ability to meet target constraints. The constraints identified another set of certainty companies. These were companies whose R&D measures were so large that the company had to be included in order to satisfy the constraints for the industry groupings. These are referred to as "analytical" certainties.

Every company in the frame was subjected independently to a selection procedure based upon its assigned probability. Thus, a given company's selection (or nonselection) did not depend on any sampling results occurring before or after it was subject to sampling. This implied that the resulting sample size could not be fixed. If the same sampling scheme were repeated, a different sample size would, likely result. The average or expected sample size over repeated samplings was the desired or specified total sample size. Since the variability of the sample size at the total universe level was small, the actual sample size did not deviate far from the desired value. At lower levels, this was not necessarily the case. Verification tabs were developed to allow the quality of the sample to be evaluated at the industry grouping levels. Samples not meeting requirements were reselected.

Another aspect of the sample design affected the expected size of the sample. This was the imposition of a minimum probability rule. The pps design resulted in the very small sampling units receiving extremely small probabilities of selection, and consequently, large sample weights if they were selected. If such a company reported an R&D value inconsistent with what was anticipated, it could have a damaging affect on the survey estimates. The minimum probability rule lessened the impact of this type of situation. The rule specified that the probability of selection could not be less than a predetermined value--thus putting a cap on the maximum weight a company could receive. Any probability found less than this predetermined value was set to this value. However, whenever probabilities were raised, the expected sample size increased. This increase was not necessarily trivial for the R&D frame, which was large (about 450,000 companies) and heavily skewed to small companies. This rule accounted for most of the difference between the specified and actual sample size.

Unique Problems/Issues: A second "response analysis" study was conducted by the Census Bureau and the NSF between November 1982 and April 1983 for the 1981 R&D survey. The reasons for conducting this response evaluation were to identify questions or definitions that created reporting problems and unnecessary reporting burden, to gather information for improving the survey form, to examine the quality of the data, to determine whether or not to support a request that all reporting in the survey should be mandatory, and to prepare for the OMB clearance of the survey for 1984.

The Census Bureau personnel visited 96 companies, including the largest R&D performers in the major industry groupings. These companies were reporters in the 1981 survey.

The most frequent criticisms related to definitions for scientists and engineers and the distinction between basic, applied, and development R&D expenditures. Many companies found these definitions difficult to understand. Another finding was that poor response to three questions added to the RD-1 form in 1979 was related to the cost incurred by companies in obtaining these data and the data's perceived lack of importance by the companies. The three questions added were: (1) (Item 14) Product versus Applied R&D; (2) (Item 15) Short versus Long-term R&D Costs; and (3) (Item 16) Total Company Funds (to meet regulations of certain agencies, etc.). Item 16 last was collected in 1983, while Items 14 and 15 last were collected in the 1987 survey. From this evaluation, it was recommended that "Computer Systems" development be added to the list of product fields. This category appeared most frequently under "Other" in the product field question. Another recommendation was that data on basic, applied, and development R&D by product group be collected every year.

Comparing long forms from 1983 and 1985, "Computer Sciences" was added to the Fields of Basic Research item (Item 7) in 1985. This apparently was a result of a recommendation from the response analysis conducted in 1983. Also in 1985, the "Atomic Energy Devices" product group was deleted from the long form, while "Communications Equipment" was split out from "Electronic Components," and "Professional and Scientific Instruments" was split into "Scientific and Mechanical Measuring Instruments" and "Optical, Surgical, Photographic, and Other Instruments." For these later changes, however, we are not sure if they were made based on results from the response analysis.

1987-1991

Scope: Coverage changes were made for this survey period as the NSF reduced the list of nonmanufacturing industries subject to sampling by eliminating those assumed to have little or no R&D activity. Also, variable employee cutoffs were used for most in-scope industry groups which caused small companies, not expected to have R&D, to be excluded from the frame. The NSF provided a list of companies to be added with certainty. Small single units (below 5 employees) that were part of the Census Bureau's administrative record universe were eliminated from scope for reasons stated earlier.

Even though a small amount of R&D activity would be missed as a result of these changes, it was determined that the efficiencies resulting from a much smaller frame were considerable. The result was that the total frame size dropped from about 450,000 companies in the 1981 sampling operation to about 154,000 companies in 1987. With the sampling frame reduced by such a significant amount and the size of the sample allowed to increase somewhat, improved national estimates of total R&D expenditures and employment were expected.

The alternating use of long-form and short-form versions of the RD-1 form continued during this period. Companies designated to receive Form RD-1A the first year of the panel were imputed (as discussed earlier) in subsequent years unless they were converted to large-company status.

Frame Sources: For the first time, the SSEL file served as the frame source for both multiunit and single-unit companies in the 1987 sampling. From the SSEL, a sample of approximately 13,900 manufacturing and nonmanufacturing companies was selected. As in previous periods, lists of large R&D contractors provided by DOD and NASA were reviewed and added to the panel as certainties.

Classification: The classification of multiunit companies in this period followed closely the procedure used in 1981. Based on SSEL payroll information and establishment SIC-4 codes, the largest of the 10 economic sectors was determined. Within that sector, the largest SIC-3 code, also based on payroll, was found. This code became the company code and determined in which industry grouping (recode) the company was classified. For some companies, the NSF converted this code, if different, to the code assigned in the previous period. The number of published industry groupings remained at 26 throughout this period.

As in the two previous sampling periods, an independent prior-period estimate (1986) was collected in the first year of the new panel. This estimate formed the basis for revising the 1981-1986 historical series using the wedging procedure. Wedging revised the intermediate years to account for coverage loss of the old panel over time and for coding changes that would have occurred had old panel codes not been frozen. The 1981 original estimate served as one link since it was based on the first year of the old panel and suffered no coverage or coding change problems. The 1986 revised estimate from the new panel served as the second link for the same reasons.

The estimates for 1987-1991 were a continuation of the wedged series. Since company codes remained frozen during this span, these series would be revised once an independent estimate for 1991 was formed from the new panel drawn for survey year 1992.

Sample Design: The basic pps sample design begun with the 1981 sample selection was continued. There was, however, a minor refinement made to the imputation of the R&D measure of size for companies that did not report a value in the previous panel. Imputation factors were derived separately for single units and multiunits. All companies with total employment of 500 or more were brought into the panel as predetermined certainties. All active companies from the old panel with an R&D value of $1 million or more also were included as predetermined certainties. Relative standard error constraints on the frame R&D totals again were assigned as a basis for allocating the sample across industry groupings. The minimum probability rule was employed again as a means of lessening the impact of small companies reporting R&D far above what was indicated by their allocated measure of size.

Unique Problems/Issues: A third "response analysis" was conducted for the 1987 R&D survey starting in mid-1988. Again, it was a combined effort by the Census Bureau and the NSF. The purpose of this response analysis was similar to that of the previous two. Seventy-six large R&D-performing companies were visited by Census Bureau personnel. Of interest was the recognition that reporting for the number of scientists and engineers was inconsistent and was subject to interpretation and records available for making estimates. This problem was similar in nature to the one discovered in the previous response analysis. The recommendation was to review the need for precise specification of scientists and engineers. Also of interest was the recommendation to evaluate the need for additional product lines for applied R&D. It seemed that respondents had a difficult time classifying "Computer Software" and related items among the current product groups.

Comparing long forms (RD-1) for 1987 and 1989, Items 11 and 12 (Product versus Process Applied R&D and Long versus Short-term R&D Costs, respectively) last were collected in the 1987 survey. We found no other significant changes in the long form.

In 1988, the NSF observed that 1986 Federal R&D support data were not calculated properly. Data from the merger of two companies were not correctly processed, which resulted in an overcount of $356 million. This was corrected for the 1989 publication. Another adjustment in 1988 was made to correct for rounding errors of numerous small firms that occurred in 1987. Due to the enlarged sample in 1987 (13,900 in 1987, up from 11,500 in 1981), many firms were receiving this form for the first time, and were not used to providing answers rounded to the nearest thousand. The following year, when these companies received the form with the imprinted prior-year data, they corrected the data, resulting in an overall downward revision to the 1987 data of over $1,717 million. The revision was reflected in the 1988 publication.

In 1989, many new issues were addressed: (1) Several small firms reported they had no R&D program in 1988 when, in fact, they reported R&D expenditures for the 1987 survey. After verifying that they truly had no R&D in 1987 and 1988, the Census Bureau removed the data from the 1987 estimates during the survey processing in 1988. (2) There were also several similar problems that occurred in the 1989 data collection and processing. (3) After reviewing the 1989 estimates, an error was revealed in the revision procedure used for firms discovered to be out-of-scope in 1988.

In response to a request from OMB, the NSF asked the Census Bureau to conduct a test using the 1990 survey to determine if combining both mandatory and voluntary items on one questionnaire influenced response rates. The 1990 sample was divided into two panels of roughly equal size. The "mandatory" panel reported the four normal mandatory items (total R&D expenditures, Federal R&D funds, net sales, and total employment) with the remaining items being voluntary, while the "voluntary" panel reported all items on a voluntary basis. The response rates for the "mandatory" and "voluntary" panels were 89 percent and 69 percent, respectively. The overall survey response rate dropped to 80 percent from levels of 88 percent in 1989 and 89 percent in 1988.

1992-1994

Introduction: Sample selection activities for the 1992 survey year marked the beginning of annual sampling for the R&D survey. Several benefits were realized from this change. Foremost was the fact that no panel deterioration occurred as with the continuing panels. With the identification each year of the R&D universe frame, a representative sample from that frame was selected. The continual updating of the SSEL source file ensured that mergers, acquisitions, births, deaths, etc., were reflected each year in the sample frame.

Secondly, companies were coded on a yearly basis. This allowed, regardless of the amount of shifting of companies from one industry grouping to another, an appropriate allocation of the sample each year to ensure adequate representation of industry groupings. With the continuing panel, if selected companies had been allowed to change, industry control achieved with the initial sampling would have been lost. In extreme (but unlikely) scenarios, all selected companies of a given industry grouping could have switched out of the grouping while no selected companies switched into the grouping. The result would have been a panel with no representation in that industry grouping. This loss of control was, we believe, the primary reason for freezing codes in the past.

The two factors mentioned above, loss of coverage over time and the assumed changing of company codes over time, were precisely the shortcomings that the wedging methodology addressed. Thus, with annual sampling, it was no longer required to wedge the historical series. The data series 1987-1991 were the last series to be so adjusted.

In this report, we have chosen to discuss these 3 years together. This is because the survey design activities related to these 3 years are virtually the same.

Scope: In 1992, the survey scope was expanded beyond recent designs to include nonmanufacturing industries generally regarded in the past as not having significant R&D activity. Enough activity was observed in 1992 that these industries were retained for both 1993 and 1994. The added industries included SIC groupings 07-09, 51-52, 55, 57-59, 61, 64-65, 67, 701, 75, 76, 79, 80 (except 806 and 807), 81, 872, 874, and 899. Employment size cutoffs, used in the 1987 sampling operation, were not used in any of these years. The absence of documentation explaining the basis for the previous cutoffs, and the uneasy feeling that important R&D activity might be missed, led to this decision. Analysis of small companies selected during these years is on-going, and this work may ultimately lead to a reconsideration of employment cutoffs. Thus, the only companies excluded because of size considerations continued to be single-unit cases which were treated as administrative record companies in the economic censuses. These sampling scope changes dramatically increased the size of the universe frame. In all 3 years, roughly 1.8 million companies were in scope. Roughly half of the increase from the 1987 universe size was due to the new industries and half to the elimination of the employment cutoffs.

Improved identification of not-for-profit enterprises took place during these years. Nonprofits always had been defined as not in scope of the R&D survey, but during most of the years of the survey, they could not reliably be identified and eliminated from the frame sources. As a result, a small number of nonprofits would be sampled. This was further complicated by the fact that some enterprises had both nonprofit and profit activity. The SSEL provides a flag for each establishment indicating whether or not the establishment is nonprofit. This flag was used to eliminate nonprofit establishments from consideration in company classification; and if all its establishments were nonprofit, it led to the removal of the company from the frame. There is undoubtedly some nonprofit activity that has not been identified on the SSEL, but it is believed we have certainly improved the frame by use of this identifier.

Alternating use of a long-form and short-form version of the RD-1 form (now labeled RD-1L and RD-1S, respectively) for large companies continued for these years. Smaller companies (generally companies new to the panel or overlap companies not reporting large amounts of R&D) received form RD-1A.

Frame Sources: The SSEL was the prime frame source for both multiunit and single-unit companies for all 3 years. With the increases observed in the frame sizes, sample sizes for these years were nearly doubled from the previous period and ranged between about 23,400 and 24,100 companies. In 1992 and 1993, several sources outside the Census Bureau were utilized to identify companies likely to conduct R&D. The source material for most of these outside sources were the 10K and 10Q reports that are filed with the Securities and Exchange Commission by publicly-owned companies. These companies, when they could be verified as residing on the SSEL, were added to the panel as predetermined certainties. Companies that could not be found on the SSEL were not added in order to avoid possible duplication; for example, it was assumed that they were on the file but could not be matched. The outside sources included Business Week's R&D Score Board (a list of the top 100 companies based on research and development expenditures), a CD-ROM of Moody's information on approximately 10,000 firms, Inside R&D Weekly, CompuStat Database, and others. These sources were not used in the 1994 survey design because it was felt that, in addition to the companies identified for the 1993 survey, few new companies if any could be found for the 1994 survey without expending a great deal of time and energy.

Classification: A slight modification to the classification of multiunit companies occurred beginning with the 1992 survey. A hierarchical approach was followed in assigning a company SIC-3 code. The first stage in this assignment did not change. Establishment payroll data were summed to determine the largest of the 10 major economic sectors for each company. Nonprofit activity, as mentioned above, was eliminated when it could be identified. Thereafter, the largest SIC-2 (payroll based) within the largest sector was determined, and then the largest SIC-3 within this SIC-2. This SIC-3 code became the company code and determined in which industry grouping (recode) the company was classified. There was no overriding of the assigned recode if it differed from the previously assigned code as happened for a few companies in the 1987 sampling operation. For 1994, a small number of companies was identified that could not be coded to an SIC-3 code because of incomplete coding in the SSEL. A small sample of these cases was included in the 1994 panel (as described in the "Sample Design" section of this period). The number of published industry groupings remained at 26 for these years.

Wedging was performed for the last time to revise the 1987-1991 data series. The independent 1991 estimate collected during the 1992 survey year served as the ending link point while the original 1987 survey estimate served as the beginning link point. There were differences between the new panel and the old panel which were not addressed by wedging and became topics of concern. For example, the 1987 SIC manual was the basis for coding the new sample, but in the 1987 sample selection, all coding was based on the 1982 SIC coding structure. This, along with the fact that the change in methodology for assigning codes was changed somewhat between the two periods and the survey scope was changed significantly, introduced discontinuities that were not properly addressed by wedging.

Although it was not possible to adjust either sample to account for each of these differences, one adjustment was made to measure the effect of the change in the classification methodology. Companies in the 1992 panel were recoded based on the previous methodology and new estimates were produced. It was found that the number of companies receiving different classifications was small, but the impact on R&D expenditures for a few industry groupings was significant. Revised estimates for 1991, based upon the reclassification of the new panel, were generated. For research purposes, these revised estimates served as a new link point for a second wedging operation. This wedged series was referred to as the "analytical" series. All steps generating the analytical series and the classification analysis are extensively documented and are referenced later in this report. The analytical series did not replace the original wedged series in any historical summary.

The revisions to the data series 1987-1991 were the last to be made using the wedging methodology. However, yearly revisions continued to be made on the basis of changes made to prior-period data by respondents.

Sample Design: The basic pps design remained in place for these years, and modifications were gradually introduced to enhance design implementation. For example, scaling the sampling frame measures of size was introduced beginning in 1992 so that the summed totals of the measures (by recode) closely approximated the most recent estimated totals for R&D expenditures. By preserving the observed relationships between recode totals, we believe better allocation of the sample among recodes will be attained. This procedure still would not result in a distribution of frame R&D values that exactly resembles the universe of true values, so it was not expected that relative standard error constraints, specified during the sampling operation, would be necessarily satisfied by the actual survey estimates. The purpose of assigning constraints was to establish a relative importance for the sampling strata, and, therefore, to fix the allocation of the sample while attempting to achieve a target precision on our estimates.

Sampling strata were redefined for these years. In the 1981 and 1987 sampling operations, relative standard error constraints were imposed for each of the 26 recodes. The recode levels corresponded to the levels at which the R&D estimates had been historically published and the target reliability constraints attempted to control precision of the estimates at these levels.

The NSF expressed a desire to publish at less-aggregated levels beginning in 1992, and after much discussion the 26 industry groupings were disaggregated into 165 groupings for sampling purposes. Each of the 165 strata were defined by one or more SIC-3 codes. Relative standard error constraints were imposed for each of the 165 strata, and the allocation of the 1992 sample was determined by these constraints. This stratification was repeated in both 1993 and 1994. Despite this change, the 1992 and 1993 data continued to be published at the recode level. The Census Bureau will evaluate whether acceptable estimates at the lower levels, or at some consolidation of these levels, could be made. This has important implications since sampling at these lower levels does not produce the most efficient sample if we are only to prepare estimates for the higher level (recode) aggregates.

A more fundamental change in the sample design was made for survey year 1994. A Census Bureau study of the 1992 survey results observed that a disproportionate number of small companies, usually reporting little or no R&D activity, was included in the frame. This was caused by the tremendous influx of small companies that occurred with the expanded scope of the survey in 1992. This increase in the proportion of small companies in the frame caused a higher proportion of small companies to appear in our panel. This increase also resulted in a greater application of the minimum probability rule. The cumulative increase of the probabilities resulting from this rule raised the expected sample size by several thousand. To address this, the 1994 sampling frame was split into two partitions. For each of the 165 sampling strata, a payroll cut-off value was found which ensured that companies above the cutoff accounted for at least 90 percent of the total stratum payroll. All predetermined certainties were included in the above cutoff partition even though their payroll value may have been below the cutoff value. Across all strata, about 640,000 companies were in the "large" partition and about 1.2 million were in the "small" partition.

The usual pps sampling scheme with a minimum probability rule was used for the large partition. The minimum probability applied in the 1994 survey was 0.002. For the two previous survey years, the minimum probability was set to 0.00101. Based on variance criteria established by the survey staff, approximately 18,000 companies were chosen. For the small partition, a more efficient simple random sample design was used. Simple random sampling, with a fixed sample size, was more efficient than pps sampling for this class of companies since there was little variability in the measures of size. A sample of about 5,200 was selected initially.

This sample included companies that could not be assigned an SIC-3 code. These companies were treated as a separate stratum in the small partition and were sampled separately. The total sample of 5,200 was proportionately allocated among the strata based on payroll. The sample size for some strata was adjusted, however, if the allocated size resulted in a sampling weight exceeding 400 for selected cases. The sample of 5,200 was roughly less than half of what was sampled from this class in 1992.

Unique Problems/Issues and Current Status of Work: During this period, a total of three separate research projects were entered into by the Census Bureau and the NSF for improving the quality and timeliness of the survey. Some of this work is on-going, but a number of final reports have been issued, some of which have been referenced in earlier sections.

First Project: The first project included three tasks.

Task 1: Selecting a new sample for the 1992 survey.

Task 1 has been completed and the following reports discuss the activities described under this task:

1. Greenberg, 1993, "REVISION - Creation of the Sampling Frame for the Survey of Industrial Research and Development (R&D) Dated March 29, 1993," and Greenberg, 1993, "Creation of the Sampling Frame for the Survey of Industrial Research and Development (R&D) Dated April 2, 1993," internal memoranda

2. Champion, 1993, "Description of 1992 Survey of Industrial Research and Development Sample Design," document sent to the NSF

The first document discusses the creation of the sample frame, and the second is an in-depth discussion of the 1992 sample design.

Task 2: Designing a computer processing system to process the short and long-form versions of the R&D survey form.

Work on Task 2 is on-going, and new edit tests have been identified to be incorporated for 1994 survey processing.

Task 3: Conducting a telephone follow-up survey of nonrespondents to the 1991 survey.

For Task 3, the survey has been completed and a final report is in draft. Several recommendations for techniques to improve response for both large and small companies were made. The report also provided recommendations for improving response to detail items for companies providing only the mandatory data.

Second Project: The second project consisted of five tasks.

Task 1: Studying methods used to follow up nonrespondents in the R&D survey and compensate for nonresponse, including imputation, and analyzing the nonrespondents to determine where there may be major problems in the R&D survey.

A draft report describing the follow-up procedures used for the 1992 survey has been written. Imputation procedures are described in technical specifications, but new documentation will be prepared. A project to develop tabulations of 1992 survey data to provide information on response rates by cross-classified categories is beginning.

Task 2: Evaluating survey quality by studying the sources, control, and measurement of nonsampling errors, and proposing follow-up studies where more work is needed on nonsampling errors. Three reports were issued on this task.

1. Bond, 1994, "Documentation of Nonsampling Error Issues in the Survey of Industrial Research and Development," Statistical Research Report Series # RR94/03

2. Bond, 1994, "An Evaluation of Imputation Methods for the Survey of Industrial Research and Development," Economic Statistical Methods Division (ESMD) Report Series # ESMD-9404

3. Bond, 1994, "A Study of Processing Errors in the Survey of Industrial Research and Development," ESMD Report Series # ESMD-9403

The first report documents the current state of knowledge on nonsampling errors in the R&D survey including what has been done to control and measure nonsampling errors. The second report contains results from an evaluation of the imputation methods used for companies mailed form RD-1A. These are the smaller companies that have little or no R&D activity. This report concludes that the current imputation methodology is adequate for partial respondents who report some, but not all, item data <u>if</u> the auxiliary data needed for imputation are present. It recommends that backup procedures be developed for cases where auxiliary data are not available. It also recommends that the assumption that unit nonrespondents are R&D nonperformers be investigated. A study of processing errors in the survey is documented in the third report. This study found little error introduced in keying, but, among several recommendations, this report suggests that unnecessary keying of reported zeros be eliminated.

Task 3: Investigating the scope, frame construction, and sample design of the R&D survey. Sample selection activities for the 1993 survey were conducted under this task and have been completed. Several reports, including this historical perspective, have been issued as part of this task.

1. Kusch and Ricciardi, 1995, "Design of the Survey of Industrial Research and Development: A Historical Perspective," Manufacturing and Construction Division Working Paper Series # Census Bureau/MCD/WP-95/01

2. Ricciardi, 1995, "Examining Variances and Standard Errors for Companies with Total Employment (TE) < 1000 from the 1992 R&D Data File," internal memorandum

3. Tulp, 1994, "Evaluation of Total Employment Cutoffs in the Survey of Industrial Research and Development (R&D)," internal memorandum

4. Kusch and Ricciardi, 1994, "Comparison of Company Coding Between 1992 and 1993 for the Research and Development (R&D) Survey," internal memorandum

5. Greenberg, 1994, "Creation of the Sampling Frame for the Survey of Industrial Research and Development (R&D)" and Greenberg, 1994, "REVISION - Creation of the Sampling Frame for the Survey of Industrial Research and Development (R&D) Dated March 14, 1994," internal memoranda

The first report documents the survey design history of the R&D survey from 1957-1994. Report 2 examines the increase in variance observed between 1991 and 1992. While some increase in variance was expected because of the dramatic increase in the size of the in-scope universe, the report concludes that this increase was exacerbated by a disproportionate number of small companies being selected from the universe of small companies. The minimum probability rule, most likely, was the chief reason for this occurrence. Results from this study were the basis for the change in the 1994 sample design in which the small companies were sampled separately under a simple random sample design. The third report documents the evaluation of R&D reported in survey year 1992 by companies that would have been excluded from the survey if the employment cutoffs used in the 1987 sample design had been retained. The results showed that for some recodes and sampling strata, a nontrivial amount of R&D activity occurred, but they also suggested that there are areas where cutoffs might be reasonable. This analysis will be repeated on 1993 survey data to see if consistent results are obtained.

Report 4 examines changes in company coding in the frame universes that occurred between 1992 and 1993. This report found that while the number of companies changing recodes was small, the effect on R&D estimated totals was sometimes meaningful. This study will be conducted for subsequent universe frames to see if consistent patterns develop which might suggest that rules be implemented to "resist" coding changes. The final document discusses the creation of the sample frame for the 1993 R&D survey.

Additional reports will be forthcoming under this task. These include documentation of work evaluating the quality of estimates resulting for the 165 sample strata used in the sample designs for 1992-1994. This analysis will consider what increase in sample size is required to achieve reliable estimates at these levels or at various consolidations of these levels, and will examine the effects on total sample size of various alternative stratification schemes. A report also will be prepared on comparisons of different methods for deriving the R&D measures of size used in the pps design.

Task 4: Phase I: Evaluating survey forms and letters, including interviews of respondents to find out how they interpret survey questions and concepts. The following report has been issued under Phase I.

1. DeMaio and Davis, 1993, "Review of the Report Form, Instruction Booklet and Letters for the Survey of Industrial Research and Development," document sent to the NSF

A research study (Phase II), in which respondents evaluated the survey form, was conducted in calendar year 1994. Analysis of interviews conducted by phone and in person is almost complete. A final report will be forthcoming under Phase II.

The Phase I report discusses results from the review of the RD-1L form and instruction booklet, as well as recommendations for changes to the letters sent to respondents.

Task 5: Evaluating recommendations from tasks 1 through 4.

While some changes have been introduced based on some of the completed work, the basic charge of this task awaits full completion of tasks 1 through 4.

Third Project: The third Census Bureau-NSF project was composed of three tasks.

Task 1: Evaluating electronic data collection needs for the R&D survey panel members that report annually. Work on this task has been initiated. A total of 200 companies have been identified to test the use of an electronic questionnaire. These companies are split into two panels--one to test the electronic questionnaire and one to serve as a control group.

Task 2: Providing an automated system for transmitting copies of the survey reporting form to small R&D performers via fax. This system is in place.

Task 3: Conducting an evaluation of existing and proposed data items for the R&D survey. Two reports have been issued under this task, but priority given to 1993 survey processing and development of the electronic questionnaire prevented further analysis at that time.

1. Champion and Capps, 1993, "Evaluation of Proposed Changes to the Survey of Industrial Research and Development," document sent to the NSF

2. Champion, 1993, "Evaluation of RD-1A Test Form of the Survey of Industrial Research and Development," document sent to the NSF

The first report documents the evaluation of proposed changes to the R&D survey. The purpose of the evaluation was to interview companies in the current survey panel about their understanding of the proposed changes, their views on the clarity of definitions given, their ability to report the data, and their recommendations for improvements to the survey. The results of the evaluation were based on phone interviews and personal interviews with 84 companies. Of those interviewed, 49 companies were in the top 200 R&D performers. One of the more interesting results from this evaluation was the recommendation to add five new categories to the list of products in Item 8. These five new categories were: (1) paper, (2) leather, (3) lumber, (4) wood products, and (5) computer software. The second report summarizes results from the evaluation of the RD-1A form for the 1992 R&D survey. The purpose of the evaluation was to interview companies that are not in the current survey and determine their understanding of the form and instructions, the clarity of the definitions, their ability to report the data, and any recommendations.

Nine companies were visited and asked a series of questions about the instructions and design of the form. The response was very positive and the firms appreciated the opportunity to participate in the evaluation. The majority of the respondents understood the survey and its concepts, and felt that the reporting burden was at a minimum.

1992 Wedging Analysis: In addition to work produced under these agreements, a considerable amount of analysis was produced related to 1992 wedging of 1991 data. As was noted earlier, several differences between the 1987 design and the 1992 design existed which wedging did not address properly. Thus, although the usual wedging operation was performed, a second "analytical" wedging also was done for research purposes only. This was based upon a second 1991 estimate that was derived after the 1992 panel companies were recoded using the methodology of 1987. A series of reports describing the wedging operation in general, the differences between the two years, and the reclassification, were produced.

1. Tulp, 1994, "Reclassification of Companies in the 1992 Survey of Industrial Research and Development (R&D) for the Generation of the Analytical Series," internal memorandum

2. Kusch, 1994, "Wedging Considerations for the 1992 Research and Development (R&D) Survey," internal memorandum

3. Tulp, 1993, "Effects of the 1987 SIC Revision on Company Classification in the Survey of Industrial Research and Development (R&D)," internal memorandum

The first report documents the activities involved in reclassifying the 1992 R&D sample so that the "analytical" series could be generated. The second report describes the wedging methodology and its properties, and discusses the official and analytical series and their limitations. The third report documents research done on the possible effects of the 1987 SIC revision on company classification in the R&D survey. The results suggested that overall, the 1987 SIC revision did not largely affect the classification of companies in the R&D survey. There were a few recodes that did experience significant changes, however, the company recode changes represented only 1.1 percent of the total number of companies (150,568) included in the analysis.

These reports form a comprehensive summary of the wedging operation and the analysis which went into producing the analytical tables.

Note on References: The majority of the information gathered for the historical document came from the "Technical Notes" section of past NSF publications for the R&D survey (or example, Funds for Research and Development in Industry, National Science Foundation, Washington). The years of the publications spanned from 1957 through 1990. The 1990 publication was in draft form when this document was written. Additional references used to supplement this information were recent OMB packages and Census Bureau specifications. The OMB packages described many activities of the R&D survey in general, while the specifications detailed many of the methodologies employed in the R&D survey.

Acknowledgements: The authors want to thank Brian Greenberg, Stacey Cole, Elinor Champion, and Dan Tulp at the Census Bureau and Ray Wolfe and John Gawalt at the NSF for their valuable contributions and suggestions to and comments on earlier drafts.

1957-1961

Scope

- •1957 - Covered all manufacturing and selected nonmanufacturing industries. Trade associations and agricultural cooperatives were excluded.
- •1958 - Covered a representative sample of companies in virtually all nonmanufacturing industries. Federal contract research centers were included for the first time. The railroad industry and nonmanufacturing companies with less than 50 employees were excluded.

Frame Source

- •1957 - Consisted of the Annual Survey of Manufactures (ASM) from the Census Bureau and a list from the Bureau of Old-Age and Survivors Insurance (BOASI). The frame was supplemented in nonsurvey years by business births and from an annual Department of Defense (DOD) list of the largest R&D contractors.
- •1958 - Multiunit companies with more than 1000 employees not represented in the current panel were added with certainty.

Classification

- •1957 - There were 18 industry groupings (16 manufacturing and 2 nonmanufacturing), but were expanded to 22 in 1958 (21 manufacturing and 1 nonmanufacturing). The next change took place in 1960 when stone, glass, and clay was included in the "other manufacturing" category. In 1961, stone, glass, and clay was treated as a separate industry grouping again.
- • "" - Classification for manufacturing companies was based first on employment taken from the ASM. Then within manufacturing, a code was assigned on the basis of value added. Classification of nonmanufacturing companies was taken from BOASI industry codes.
- •1958-9 - A slight change occurred by which a manufacturing company was classified into an industry on the basis of its product class shipments.
- •1960 - Industry codes were based on 1957 Standard Industrial Classification (SIC) manual. Prior to 1960, these codes were based on 1947-49 SIC manuals.

Sample Design

- • The sample design for these years used a stratified random sampling scheme.
- •1957-61 The manufacturing sample was based primarily on the 1956 ASM.
- •1958 - A supplemental sample was drawn to measure the R&D activities in small manufacturing firms. The "shuttle form" was first implemented. This allowed companies to review prior year data and make any necessary corrections to the questionnaire.
- •1958-61 The nonmanufacturing samples were drawn from the 1956 BOASI records.

Sample Characteristics

	1957	1958[a]	1959	1960	1961
•Size	6800	6300	6700	6700	6700
•Certainties	2000	1200	1900	1900	1800
•Cert. Criteria (Emp. >)	2500	1000	1000	1000	1000

1957-1961 (cont'd)

Unique Problems/Issues

- •1957 - The Census Bureau requested 1956 data as part of its survey to evaluate comparability of its estimates with those of the Bureau of Labor Statistics (BLS). The BLS estimates were 4 percent higher than the Census' for 1956. Some industry percentage differences were even higher than for the overall total. Differences mainly due to the way companies reported for the two surveys, and to variations in the methodologies used by the two bureaus. Due to differences mentioned above, is was not possible to present consistent trend data for individual industries for 1953-57. For overall comparisons, however, the National Science Foundation (NSF) was able to adjust the figures and provide a basis for time series for 1953-57.
- •1958 - The share of R&D performance that was financed by the Federal Government may be somewhat understated in certain industries, because some company-financed R&D is indirectly financed through overhead payments under Federal contracts. It was difficult for companies to separate company-financed R&D from R&D financed through overhead charges. Thirdly, in 1958, a supplemental sample was used to address the problem of identifying R&D in small firms.

1962-1966

Scope

- • No changes in scope occurred in this survey period.

Frame Source

- • The primary frame sources continued to be the Bureau's ASM files and the BOASI files. Panels were still updated in nonsample years by a list of the largest R&D contractors from DOD.

Classification

- •1962 - The number of published industry groupings remained at 22 for this period.
- •1965 - From 1963-65, industry classifications were based on data reported in 1963 Economic Censuses. For 1962 and earlier, industry classifications were based on data reported in 1958 Economic Censuses. Between 1958-63 some companies changed SIC codes and the industry data for the affected years (1959-62) were estimated to have changed at a constant rate of 20 percent per year. Accordingly, the data for the industry in which a company had been classified in 1958 were deflated by 20 percent per year for each of the affected years. Similarly, the data for the industry in which a company was classified in 1963 were inflated by 20 percent per year for each of the intervening years.

Sample Design

- • No major changes occurred for this sampling operation. Stratified random sampling continued to be used with the stratification by industry grouping and employment size.

1962-1966 (cont'd)

Sample Characteristics

	1962	1963	1964	1965	1966
•Size	7000	7000	7000	7000	7000
•Certainties	1800	1800	1800	1700	1700
•Cert. Criteria (Emp. >)	1000	1000	1000	1000	1000

Unique Problems/Issues

- •1962 - The results of two questionnaire items, "Forward Budgeting of Company Funds for Research and Development," and "Capital Expenditures for Research and Development" were not published because response to these items was either lacking or suspect.
- • "" - The problem of identifying R&D activity in small companies was confirmed again when the new sample was drawn in 1962. Estimates obtained in previous years for this small company class (less than 100 employees) compared to the current year have shown to vary significantly. In 1957, it was estimated that there were 6800 small companies performing R&D compared to 12,500 in the 1958 supplemental sample and 12,000 in the 1962 sample.

1967-1970

Scope

- • The scope of the R&D survey for this period did not change.

Frame Source

- •1967 - The frame sources used in the 1967 sampling operation included: (1) the 1963 Census Enterprise Statistics file for multiunit manufacturing companies, (2) the 1963 Census of Manufactures file for single-unit manufacturing companies, and (3) the 1966 Social Security Administration (SSA) records for nonmanufacturing companies. The panel was further supplemented in nonsample years from DOD lists of R&D contractors and also from a list of R&D contractors provided by the National Aeronautics and Space Administration (NASA).

Classification

- •1967 - The number of industry groupings increased to 23 for survey year 1967 (the radio and television receiving equipment industry was split out from other electrical).
- • "" - The only significant change that took place in this period was that classifications were now based on the 1967 SIC manual. Prior to 1967, the 1957 SIC manual was used.

Sample Design

- • The basic sample design remained unchanged for this sampling operation. Stratified sampling continued to be used.

Sample Characteristics

	1967	1968	1969	1970
•Size	8000	8000	8000	8000
•Certainties	1800	1800	1800	1800
•Cert. Criteria (Emp. >)	1000	1000	1000	1000

1967-1970 (cont'd)

Unique Problems/Issues

•1967 - Data collected from the 1967 sample suggested again, as in 1958 and 1962, that there was a problem in identifying R&D in small companies (employment less than 100). In 1967, it was estimated that 10,000 small companies conducted R&D compared to about 12,000 in 1962, and about 12,500 in 1958 from the supplemental sample. The 1957 sample estimate of about 6800 emphasized this variability to an even higher degree.

1971-1975

Scope

• During this period no major changes in scope occurred.

Frame Source

•1971 - The frame sources used in the 1971 sampling operation, for the most part, were updated versions of the sources used in 1967. Thus, multiunit manufacturing companies were identified from the 1967 Enterprise Statistics file and single-unit companies from the 1967 Census of Manufactures file. In a slight departure from the 1967 sampling, the Enterprise file also was used as the frame source for selected nonmanufacturing industries (SIC's 7391-92, 7397, 8911). The remaining in-scope nonmanufacturing industries again were identified from SSA records. The panel continued to be updated in nonsample years from annual lists of R&D contractors provided by DOD and NASA.

Classification

•1971 - The number of industry groupings published remained at 23 for 1971 and 1972 but was increased to 25 for the remaining years. Electrical components and other transportation equipment were treated separately.

• "" - Between 1963-67, some companies changed codes and industry data for the years affected (1964-67) were estimated to have changed at a constant rate of 25 percent per year. [The similar method of deflating/inflating data used in 1965 was used here.]

Sample Design

• There were no changes in sample design from the previous survey period. A stratified random sampling scheme was still utilized to select the 1971 sample panel.

Sample Characteristics

	1971	1972	1973	1974	1975
•Size	8000	8000	8000	8000	8000
•Certainties	1800	1800	1800	1800	1800
•Cert. Criteria (Emp. >)	1000	1000	1000	1000	1000

Unique Problems/Issues

•1975 - A "response analysis" was conducted by the Census Bureau and NSF. The purpose was to discuss each questionnaire item with respondents to determine the sources used by companies, problems encountered, etc.

1976-1980

Scope

• In this survey period, the only significant change was that the detailed questionnaire was to be mailed only in the odd-numbered years after the 1977 Census. A short form containing only the most important data items was mailed in the even-numbered years to lessen the reporting burden on the respondent.

Frame Source

•1976 - A rather significant change took place in frame sources for the 1976 sampling operation. For the first time the Bureau's Standard Statistical Establishment List (SSEL) was used. The 1974 version of the SSEL was the prime frame source for all manufacturing industries and for selected nonmanufacturing industries (SIC's 49, 7391-92, 7399, 8911) for single-unit companies. Records from the SSA were used to identify single-unit companies in the remaining nonmanufacturing industries. Multiunit companies were drawn from the 1972 Enterprise Statistics file.

Classification

•1976 - Again, changes in the SIC classification system were occurring. Prior to 1976, SIC coding was based on the 1967 SIC manual. A revised manual (1972) was in place for 1976.

• "" - Between 1967-75, some companies changed codes and industry data for the years affected (1968-74) were estimated to have changed at a constant rate of 14.3 percent per year. [The similar method of deflating/inflating data used in 1971 was used here.]

• "" - Further revisions were made to the R&D data series due to prior-year (1975) data being collected in 1976 panel. This prior-year estimate was considered superior to the original 1975 estimate obtained from the old panel. Thus, panel estimates for the middle three years (1972-1974) of the old panel were adjusted for each industry grouping. Adjustments for each of the middle years were increasing proportions of the total differences observed between the two 1975 estimates. These revisions were the first of the so-called "wedging" operations performed on R&D data.

•1978 - The number of industry groupings was increased to 26. Other nonelectrical machinery was treated as a separate manufacturing grouping.

Sample Design

• There were no major changes in sample design from the previous survey period.

Sample Characteristics

	1976	1977	1978	1979	1980
•Size	11500	11500	11500	11500	11500
•Certainties	4500	4500	4500	4500	4500
•Cert. Criteria (Emp. >)	1000	1000	1000	1000	1000

Unique Problems/Issues

• There were no unique problems/issues that occurred in this survey period.

1981-1986

Scope

• There were no significant changes in scope for this period. Census single-unit companies below 5 employees (essentially the administrative record universe) were excluded.

Frame Source

•1981 - Census files continued as the primary source frames for identifying in-scope companies. In 1981, multiunit companies were identified from the 1977 Enterprise Statistics file and single-unit companies from the 1981 SSEL file. Annual lists of R&D contractors provided by DOD and NASA continued to supplement the panel throughout the survey period.

Classification

•1981 - The number of published industry groupings remained at 26 for this period.

• "" - Multiunit classifications were based on Enterprise Industrial Category (EIC) codes. Single-unit classifications in this survey period were always based on the SIC-3 code of the company.

• "" - A nonlinear algorithm, referred to as "wedging", was used to reflect changes in industry classification between 1976 and 1981. This algorithm was intended to link the two sample years (1976 and 1981) while preserving the year-to-year industry trend. A slight modification was introduced to the wedging algorithm. The new algorithm preserved, to the extent possible, the trend pattern of the original series. With the previous algorithm, the level of increase (decrease) in the revised series increased at a constant rate each year, which culminated in the last year with the full difference observed between the two independent estimates.

Sample Design

•1981 - Major changes occurred for the 1981 sampling operation. The design remained stratified but for the first time, probabilities of selection for noncertainty companies in the universe were made proportionate to a company measure of size. This is referred to as probability proportionate to size (pps) sampling. The measure of size was based on total R&D. The total R&D value was either reported or imputed.

Sample Characteristics

	1981	1982	1983	1984	1985	1986
•Size	11500	11500	11500	11500	12700	12700
•Certainties	4500	4500	4500	4500	5800	5800
•Cert. Criteria (Emp. >)	500	500	500	500	500	500

Unique Problems/Issues

•1982 - A second "reponse analysis" study was conducted by the Census Bureau and the NSF between November 1982 and April 1983 for the 1981 R&D survey. The reasons for conducting the response evaluation were to identify questions or definitions that created reporting problems and unnecessary reporting burden, to gather information for improving the survey form, to examine the quality of the data, to determine whether or not to support a request that all reporting in the survey should be mandatory, and to prepare for the Office of Management and Budget (OMB) clearance of the survey for 1984.

1987-1991

Scope

•1987 - Coverage changes were made for this survey period as the NSF narrowed the list of nonmanufacturing industries subject to sampling by eliminating those assumed to have little or no R&D activity. Also, employee cutoffs were used for each in-scope industry group which caused to drop from the frame small companies not expected to have R&D. The NSF provided a list of companies to be added with certainty. Small single units (below 5 employees) that were part of the Census Bureau's administrative record universe were eliminated from scope.

Frame Source

•1987 - For the first time, in 1987, the SSEL file served as the frame source for both multiunit and single-unit companies. As in previous periods, lists of large R&D contractors provided by DOD and NASA were reviewed and added to the panel as certainties.

Classification

•1987 - The number of published industry groupings remained at 26 during this survey period.

• "" - The classification of multiunit companies in this period followed the same procedure used in 1981.

• "" - Again, wedging was used to reflect changes in industry classification between 1981 and 1987.

Sample Design

•1987 - The basic pps sample design begun with 1981 sample selection was continued. There was, however, a minor refinement made to the imputation of the R&D measure of size for companies that did not report a value in the previous panel. Imputation factors were derived separately for single units and multiunits.

Sample Characteristics

	1987	1988	1989	1990	1991
•Size	11500	11500	11500	11500	12700
•Certainties	4500	4500	4500	4500	5800
•Cert. Criteria (Emp. >)	500	500	500	500	500

Unique Problems/Issues

•1988 - A third "response analysis" was conducted jointly by the Census Bureau and the NSF for the 1987 R&D survey starting in mid-1988. The purpose of this response analysis was simliar to that of the previous two.

• "" - An adjustment that affected the 1986 data was when the NSF discovered that the 1986 data on Federal R&D support was miscalculated. Data for a merger between two companies had not been correctly processed, resulting in an overcount of $356 million. The revised data appeared in the 1989 publication. Another adjustment was made, this time involving 1987 data. It seems many small firms in the 1987 sample made some rounding errors. These companies realized their rounding errors when they received their 1988 form with their imprinted prior-year data. When the companies revised their data, it resulted in an overall downward revision of the 1987 data of $1,717 million. Revisions are reflected in the 1988 publication.

•1989 - The following issues were addressed: (1) several small firms said they had no R&D program in 1988 when, in fact, they

1987-1991 (cont'd)

reported expenditures in the 1987 survey. After verifying that they truly had no R&D in 1987 and 1988, Census staff removed the data from the 1987 estimates during the survey processing in 1988; (2) there were also several similar problems that occurred in the 1989 data collection and processing; and (3) after reviewing the 1989 estimates, an error was revealed in the revision procedure used for firms discovered to be out-of-scope in 1988.

•1990[b] - In response to a request from the OMB, the NSF asked Census to conduct a test to determine if combining both mandatory and voluntary items on one questionnaire influences response rates. The 1990 sample was divided into two panels of roughly equal size. The "mandatory" panel reported the normal four mandatory items with the remaining items voluntary, while the "voluntary" panel reported all items on a voluntary basis. The response rates for the "mandatory" and "voluntary" panels were 89 percent and 69 percent, respectively. The overall survey response rate dropped to 80 percent from levels of 88 percent in 1989 and 89 percent in 1988.

1992-1994

Scope

•1992 - The scope was expanded in 1992 to include several new nonmanufacturing industries that were not included in the 1987 panel. Also, the scope was expanded further due to the fact that employee cutoffs used in the previous panel were not used in 1992. The result of these two actions increased the frame from roughly 150,000 in 1987 to over 1.8 million in 1992.

Frame Source

•1992-4 - The primary source for single units and multiunits in each new sample was the most current version of the SSEL. The basic frame in 1992 and 1993 was supplemented by various outside sources such as Moody's, Inside R&D Weekly, and Business Week's Score Board. The supplemental sources were used to identify predetermined certainty cases. In 1994, however, these sources were no longer used because of the time spent searching for these companies, and that very few companies were found that actually conducted R&D.

Classification

•1992 - The methodology for the assignment of a company SIC code was changed, and now followed a hierarchical approach. A company was put into an economic division based on its largest payroll. From there the SIC-2 code with the largest payroll was found within the assigned economic division. Within the assigned SIC-2 code, the largest payroll-based SIC-3 code was determined. Finally, the assigned SIC-3 code was mapped to the corresponding R&D recode.

• "" - Wedging was done for the last time for the data series 1987-1991, but due to changes in the classification system, two wedged series were produced. The original series was done as normal, while the "analytical" series was first reclassified based on the 1982 SIC coding structure, then wedged as normal.

1992-1994 (cont'd)

• "" - The number of published industry groupings remained at 26 for these years.

Sample Design

•1992 - A refinement was made which adjusted for any differences between the expected imputed R&D total and the published imputed total across recodes. The 1992 sample marked the first time that the sample was drawn annually. This eliminated birth procedures, mergers, and bought-solds that occurred but could not be detected until a new sample was selected five years later.

• "" - Sampling strata were redefined for these years. In the 1981 and 1987 sampling operations, relative standard error constraints were imposed for each of the 27 recodes (for sampling purposes, nonmanufacturing was split into two recodes, but these were combined for publication purposes). By NSF's request, the 26 industry groupings were disaggregated into 165 groupings for sampling purposes. Each of the 165 strata were defined by one or more SIC-3 codes.

•1994 - A change was made to the R&D survey to split the sampling frame into two partitions. The sample frame was partitioned across all in-scope SIC-3 codes, by determining the payroll cutoff such that the above-cutoff portion in each 3-digit industry was made up of companies representing at least 90 percent of the total payroll in that industry. PPS sampling was used on the above portion to select part of the total specified sample. From the below portion of the frame, a simple random sample was drawn such that the above and below samples together yielded the total specified sample size.

Sample Characteristics

	1992	1993	1994
•Size	23300	24000	23600
•Certainties	9700	12200	9400
•Cert. Criteria (Emp. >)	1000	1000	1000

Unique Problems/Issues

•1992-4 - During this period, three separate projects were entered into by the Census Bureau and the NSF for improving the quality and timeliness of the survey. The status of each task is listed below. **Project 1** included the following three tasks: (1) selection of a new sample for the 1992 survey (completed); (2) designing of a computer processing system to process the short and long versions of the R&D survey form (ongoing); and (3) telephone follow-up survey of nonrespondents to the 1991 survey (completed). **Project 2** consisted of the following five tasks: (1) a study of methods used to follow up R&D survey nonrespondents and measure and compensate for nonresponse, including imputation (ongoing); (2) an evaluation of the quality of the survey by studying the sources, control, and measurement of nonsampling errors, and follow-up studies where more work is needed on nonsampling errors (completed); (3) an investigation of the scope, frame construction, and sample design (under this task several reports have been issued and several more are forthcoming); (4) an evaluation of survey

forms and letters, including interviews of respondents to find out how they interpret survey questions and concepts (one report has been issued and another forthcoming); and (5) an evaluation of recommendations from tasks 1 through 4 (depends on completed work from tasks 1 through 4). The **Project 3** was composed of three tasks: (1) an evaluation of electronic data collection needs for the R&D survey panel members that report annually (ongoing); (2) providing an automated system for transmitting, via fax, copies of the survey reporting form to small R&D performers (completed); and (3) an evaluation of existing and proposed data items for the R&D survey (ongoing).

Note: [a] Indicates a supplemental sample for small firms was selected; [b] Information pertaining to 1990 was taken from a draft copy of the 1990 NSF publication.