A CAUTIONARY NOTE FROM NSF ON USING SDR DATA FOR TREND ANALYSIS:

Users should be aware that significant changes made to the SDR in the 1991-1993 period make time series analysis spanning the 1980s and 1990s data problematic. The following paper discusses work done to examine the data issues related to SDR time series analysis. Other papers in the Compendium also discuss general changes made in the early 1990s to the SDR and related SESTAT surveys.

Changes made during the early 1990s to improve the quality and utility of the SDR include:

1) Major changes in the target population definition to eliminate poorly covered populations, such as Ph.D. recipients who earned their science and engineering degrees abroad;

2) A complete overhaul of the survey instruments to reduce respondent errors, introduce more standardized concepts, and improve coverage of work force activities; and

3) Major changes in data collection methods to improve the response rate from the mid-50s to the mid-80s.

Note: The Compendium contains a separate paper on the study of SDR nonresponse bias effects. Nonresponse biases identified in that study and from research done for this paper include: some overestimation of the total size of the U.S. population of scientists and engineers, some overestimation of individuals employed in academe, and underestimation of those employed in industry. Response-related errors also were found in questions related to Federal support and for subpopulations by race.

These changes and problems have serious effects on the comparability of estimates over time. Trend analyses using the 1990s data with data from prior years must be performed very cautiously, if at all. Individuals who wish to explore such analyses are encouraged to discuss these issues with the SDR project officer, Kelly Kang (kkang@nsf.gov).

Contract No.: 9313060 and SRS-9531746
MPR Subcontract No.: NSCF-8014-98-001
MPR Reference No.: 8491-200

USING THE SURVEY OF DOCTORATE RECIPIENTS IN TIME-SERIES ANALYSES: 1989-1995

Task 2 Report

June 30, 1998

Authors:

Susan B. Mitchell
Ramal Moonesinghe
Brenda G. Cox

Submitted to:
National Science Foundation
4201 Wilson Boulevard
Room 963
Arlington, VA 22230

NSF Project Officer: Kelly Kang

Submitted by:
Mathematica Policy Research, Inc.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
(202) 484-9220
Project Director: Brenda G. Cox

Under Subcontract to:
National Research Council
2101 Constitution Avenue, NW.
Room TJ2006
Washington, DC 20418
NRC Project Officer: Peter Henderson

1. INTRODUCTION

From its inception in 1950, the National Science Foundation (NSF) has been charged "to provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources and to provide a source of information for policy formulation by other agencies of the Federal Government" (NSF Web Site 1998). The Survey of Doctorate Recipients (SDR) furnishes an important tool in accomplishing this objective. The SDR is a biennial, longitudinal survey of about 50,000 recipients of U.S.-earned doctoral degrees in science and engineering (S&E). From its initiation in 1973, the SDR has been widely used by the U.S. Congress and the federal agencies, universities and professional societies, and other organizations and individuals interested in knowing more about the nation's education, supply, and employment of doctorate recipients in S&E fields.

The SDR is a part of a larger data system NSF established to satisfy its mission—the Scientists and Engineers Statistical Data System (SESTAT). The SESTAT is a unified database recording employment, education, and other characteristics of the nation's scientists and engineers. The SESTAT contains data not only from the SDR but also data from the National Survey of College Graduates (NSCG) and the National Survey of Recent College Graduates (NSRCG). The NSCG provides data for persons who had bachelor or higher degrees in S&E or who were in an S&E occupation at the time of the Decennial Census, whereas the NSRCG provides data for persons who earned bachelor's or master's degrees in S&E after the Census. In the late 1980s, the NSF asked the Committee on National Statistics to convene an advisory panel to examine its data system and recommend changes to improve the system, which was then referred to as the Scientific and Technical Personnel Data System (Citro and Kalton 1989). The panel made numerous recommendations, the main points of which are summarized in Table 1-1. In response to the panel's recommendations, the SESTAT was extensively redesigned in the early 1990s. To improve data quality and integration, the sample design, data collection procedures, and questionnaires for each component survey were modified considerably.

Analyses across time have always been an important component of SDR's research reports. These time-series analyses typically examine estimated means, percentages, or totals generated for each survey cycle, looking for significant changes across time.[1] The major design changes that occurred in the 1990s raise the issue of whether time-series analyses have been affected by the changes and, if so, how and to what extent. In this report, we summarize the results of an investigation of the effects the design changes in the 1990s have had on time-series comparisons. We begin by presenting the SDR as it existed in 1989, and then describe the instrument and sample design changes made in 1991 and 1993 and the response rate differences that might have an impact on time-series analyses. Next, we describe a weight-adjustment procedure for the 1989 SDR that brings it into closer alignment with the sample design in the 1990s. Finally, we examine the results of using the revised weights to produce time-series estimates. Our overall conclusion is that analysts must use caution in constructing and interpreting time series that span the 1980s and 1990s.

[1] Time-series analyses differ from longitudinal analyses; the latter follow individuals over time and examine the various career paths taken. Cox et al. (1998) describe the characteristics of the SDR for longitudinal analyses.

Table 1-1. Summary of 1989 advisory panel recommendations

Establish a clear definition of the target population and criteria for eligibility.	Use other government data for subgroups not covered (those who come from abroad and those who lack a degree or have a degree in other fields).
Improve the sample design for all three surveys and for the system as a whole. Recognize that it is not cost effective to provide estimates for very small subgroups.	Find ways to improve research access to microdata from the SDR while protecting the confidentiality of individual replies.
Develop an occupation classification that is separate from major field and is based on standard occupational classifications such as the SOC.	Provide complete documentation, including a comprehensive users guide and likely magnitude of error.
Increase the response rate for all three surveys.	Publicize the data and provide support to researchers for innovative studies of science and engineering personnel.
Provide increased resources to do quality control and evaluation studies.	Establish a user panel to give input on a regular basis.
Provide increased access for data users, including improved documentation and seminars.	Communicate with users to get their input and to prepare them to take full advantage of the data.
Prepare regular profiles of characteristics of scientists and engineers.	Think of the data system as an integral whole in which all parts must work smoothly for success. Provide basic information on science and engineering personnel and detailed information on topics and subgroups of key analytic and policy interest.
Provide necessary budget and staff resources.	Adopt best survey practice in the following areas: design and evaluation; procedures for obtaining high levels of response, both through initial contact and followup; and procedures for data preparation, including developing appropriate weights, imputing missing values, and editing the data for consistency.
Strengthen the research and analysis capability of the data.	Create microdata files that permit in-depth and innovative analysis of the data.
Within framework of standard classification schemes, develop more richly detailed categorization of subgroups of scientists and engineers.	Toward the end of the decade conduct zero-based evaluation to determine what changes should be made to the system in the next decade.
Assign priority to content on work, career paths, and productivity and how they relate to training.	
Conduct large postcensal survey in 1992 that provides in-depth coverage of those college graduates employed or with training in science and engineering.	
Develop panel survey that tracks over time.	
Make modifications specific to SDR redesign coverage, survey scheduling, sample design, and wording of key items to improve comparability of the SDR data with other data in the system.	

2. SURVEY DESIGNS: 1989-1995

The Survey of Doctorate Recipients has undergone a number of changes since its inception, despite being a longitudinal survey where consistency is critical for making comparisons across time. These changes resulted from four primary sources:

- Changing policy interests;
- Changes in the survey environment;
- Advances in survey methodology; and
- Changes in project funding levels.

When the SDR was first conducted, the primary policy concern was whether the United States had enough Ph.D.s to maintain its economic competitiveness. By 1995, the focus had shifted to whether Ph.D.s might now be overproduced and underemployed. A mail survey in 1973 could achieve a 75-percent response rate without telephone followup. By 1995, such response rates were no longer possible—at least not for the SDR. Similarly, in 1973, variance estimation for complex sample surveys was machine-intensive and expensive. By 1995, such estimates were routinely produced using commercial and public-domain software. The combination of these factors produced a survey that was continually evolving and created tension between the need to maintain data comparability and the desire to improve data quality and survey operations.

The most dramatic changes occurred in the 1990s in response to recommendations made by the Committee on National Statistics (Citro and Kalton 1989). These changes, displayed in Table 2-1, affected the target population, strata, sample size, data collection mode, response rate, weighting, reference period, and questionnaire. Associated with the changes in sampling and data collection were changes in the weighting procedures (see Table 2-2). Some changes were designed to improve the quality of the survey data; other changes were made to facilitate integration of the SDR into the much larger SESTAT database. In this chapter, we first describe the SDR as it existed in 1989 and then discuss the changes that occurred in subsequent survey cycles.

A. The 1989 SDR

The 1989 SDR was the ninth survey in the biennial series begun in 1973. As with earlier surveys, its purpose was to collect data on the supply, utilization, employment, and demographic characteristics of U.S. doctoral personnel in science and engineering. Data were collected via a mail survey. The 1989 response rate was 55 percent, prompting changes to improve the response rate in the 1990s. An understanding of the statistical features of the 1989 survey is needed to understand the changes that occurred in the 1990s.

Table 2-1. Comparison of survey features: 1989 – 1995 survey of doctorate recipients

Survey Component	Survey Year			
	1989	1991	1993	1995
Target Population	Individuals who had: 1. science/engineering doctorates from U.S. institutions 2. nonscience/ nonengineering doctorates from U.S. institutions but were working as scientists and engineers 3. science/engineering doctorates from foreign institutions Doctorate degree received between January 1, 1942 and June 30, 1988	Individuals who had science/engineering doctorates from U.S. institutions Age 75 or less	Same as 1991	Same as 1991
Stratification	2,000 strata based upon degree year, sex, degree field, race, ethnicity, birthplace, and citizenship	239 strata based upon sex, degree field, and demographic group (recoded combination of disability status, race, ethnicity, and citizenship)	Same as 1991	Same as 1991 but with three substrata Substratified based on year degree received for maintenance cut
Sample Size	73,611	Reduced to 37,996	Increased to 49,228	Maintenance cut to yield 49,829
Data Collection	Mail survey No telephone followup	Mail survey Telephone followup of 65% of mail nonrespondents	Mail survey Telephone followup of all mail nonrespondents	Mail survey Telephone followup of 60% of mail nonrespondents
Response Rate*	55%	87% (63% mail only)	86% (61% mail only)	85% (62% mail only)

Table 2-1. Comparison of survey features: 1989 – 1995 survey of doctorate recipients (continued)

Survey Component	Survey Year			
	1989	1991	1993	1995
Weighting	Basic weight = inverse of sampling rate Nonresponse adjustment performed within strata No poststratification	Basic weight = inverse of sampling rate times (for telephone subsample) the mail nonrespondent subsampling rate Nonresponse adjustment made to CATI respondents only within poststrata formed by four age groups	Basic weight = inverse of sampling rate Nonresponse adjustment made to mail and CATI respondents within 1991 poststrata Small weighting classes collapsed	Basic weight = inverse of sampling rate times (for telephone subsample) the mail nonrespondent subsampling rate Nonresponse adjustment same as 1993 Collapsing same as 1993
Reference Period	February 1989 (24 months after 1987 survey)	September 1991 (31 months after 1989 survey)	Week of April 15th (20 months after 1991 survey)	Week of April 15th (24 months after 1993 survey)
Questionnaire	Question wording changes from 1987	No content changes from 1989 except for disability status	Complete redesign of survey content	Minor changes to core questions; work history and postdoc module added

* Prior to 1989, the SDR was conducted as a mail survey. Response rates experienced a steady decline from a high of 75 percent in 1973, to 64 percent in 1981, to a low of 55 percent in 1989.

Table 2-2. Comparison of weighting features: 1989-1995 survey of doctorate recipients

Weight Component	Survey Year			
	1989	1991	1993	1995
Basic Weight	Inverse of sampling rate	For mail respondents, inverse of sampling rate For CATI subsample, inverse of initial sampling rate multiplied by the inverse of CATI sampling rate	Inverse of sampling rate	For mail respondents, inverse of initial sampling rate For CATI subsample, inverse of initial sampling rate
Nonresponse Adjustment Factor (NAF)	NAF applied to mail respondents within Phase I strata	NAF applied to CATI respondents only in cells created by crossing the Phase II strata with four age groups: 1. Less than 35 2. 35-44 3. 45-64 4. 65 and up	NAF applied to mail and CATI respondents in cells created by crossing the Phase II strata with four age groups used in 1991	NAF applied to mail and CATI respondents in cells created by crossing the Phase II strata with four age groups: 1. Less than 35 2. 35-44 3. 45-64 4. 65 and up
Collapsing Rules		In cases where there were not enough CATI respondents in the cell, mail respondents that belonged to that cell were merged with CATI respondents. Under these conditions, cells were collapsed: • Cells with NAF > 2.0 • Cells with fewer than three respondents and NAF > 1.8 • Cells with no respondents	Because of the greater number of respondents in each cell compared with 1991, new guidelines were adopted for collapsing cells: • Cells with NAF > 2.0 • Cells with fewer than 20 * (1-*n*/*N*) respondents • Cells with no respondents	Nonresponse adjustment cells were collapsed using the 1991 guidelines: • Cells with NAF > 2.0 • Cells with fewer than 20 * (1-*n*/*N*) respondents • Cells with no respondents
Analysis Weight	Basic weight * NAF	For mail respondents, basic weight For CATI respondents, basic weight * NAF	Basic weight * NAF	Basic weight * NAF

1. 1989 Sampling Frame

The sampling frame for the 1989 SDR (and subsequent surveys) was compiled from the Doctorate Records File (DRF), an ongoing census of all research doctorates earned in the United States since 1920. For the 1989 survey, the science and engineering portion of the sampling frame[2] included these individuals who had academic training as scientists and engineers:

- Had earned a doctoral degree from a U.S. college or university in an S&E field; and
- Were U.S. citizens or, if non-U.S. citizens, indicated they had plans to remain in the United States after degree award; and
- Had received their degree within 42 years of the survey date.

In addition, the S&E portion of the sampling frame also contained these individuals who were employed as scientists and engineers:

- U.S. doctorates in an education or professional field employed in S&E; and
- Non-U.S. doctorates working in S&E in the United States.

The frame for individuals employed in S&E was developed in 1973 using personnel and association directories. Coverage of employed S&Es in the SDR was problematic by 1989, however, as the frame of employed S&Es was never updated.

2. 1989 Sample Design

The 1989 sample design was based upon approximately 2,000 strata formed by the intersection of cohort (or year of degree), sex, field of doctorate, race/ethnicity, and citizenship status at the time of degree award. A variable rate of sampling was applied—ranging from 2 to 100 percent—and a simple random sample was selected within each cell. The overall sample size for the 1989 SDR was 73,611.

3. 1989 Weighting

Two weights were calculated: a basic weight and an analysis weight. The basic weight was not the same as a sampling weight (the inverse of the selection probability of the unit). Rather, the basic weight was calculated as the estimated number of eligible cases in the stratum divided by the number of sampled eligible cases. This approach ignored the fact that prior restratifications may have added cases to the stratum that had been selected with different probabilities of selection. The inverse of the selection probabilities may have differed for these cases, but was not reflected in the basic weight as defined. In 1989, the nonresponse adjustment factor (NAF) for a given stratum was calculated by dividing the number of eligible cases in the stratum by the number of eligible responses in that stratum.

---

[2] From 1977 to 1995, the SDR also included doctorates in the humanities. These individuals were classified into separate strata, and their presence has no effect on the observations made in this report.

B. The 1991 SDR

The 1991 survey was conducted as a mail survey with a telephone followup of a subsample of the nonrespondents. The final weighted response rate was about 87 percent, up from 55 percent in 1989. The response rate for the mail phase alone was 63 percent.

1. 1991 Survey Content

The survey content did not change between 1989 and 1991, with the exception of the question about disability status. However, a change was made to the reference period. In 1989, respondents were asked to report for February of the survey year. In 1991, the reference month was September. Thus, between the 1989 and 1991 surveys, 31 months had elapsed, as opposed to 24 months between predecessor surveys.

The content of the 1991 questionnaire was essentially the same as that of the 1989 survey. To encourage participation, however, all survey materials were personalized with the respondent's name and address, the questionnaire was reformatted to a more "respondent friendly" design, and commercial address vendors were used to locate addresses for sample members. These changes had the effect of increasing both unit response to the mail phase (from 55 percent in 1989 to 63 percent in 1991) and item response.

2. 1991 Sampling Frame

For the 1991 survey, the sampling frame was constructed from the DRF to include individuals who:

- Had earned a doctoral degree from a U.S. college or university in an S&E field;
- Were U.S. citizens or, if non-U.S. citizens, indicated they had plans to remain in the United States after degree award; and
- Were under 76 years of age.

The coverage was poor for individuals who had earned degrees outside the United States or who held non-S&E degrees but were working in S&E. These individuals were dropped from the 1991 SDR frame.[3] Additionally, the requirements for inclusion in the frame were changed from cohort based (within 42 years of degree award) to age based (under 76 years of age). This change was made to inform policy questions regarding retirement patterns and trends.

[3] The loss of information on doctorates working as S&Es is filled by the National Survey of College Graduates (NSCG). NSCG surveys individuals with bachelor or higher degrees in S&E on April 1, 1990, and individuals with bachelor or higher degrees in other fields on April 1, 1990 who were employed as an S&E during the week of April 15, 1993. The NSCG includes individuals with degrees earned either in the U.S. or abroad.

3. 1991 Sample Design

In 1991, the decision was made to improve the response rate by telephone followup of mail respondents. Budget constraints required the reduction of the sample size from about 74,000 to 38,000 as a consequence. The resources saved as a result of the sample reduction, were redirected toward increasing the overall response rate by improving the yield of the mail survey and by using telephone interviewing to follow up a sample of nonrespondents to the mail phase. Thus, although the total sample size was reduced, its accuracy was enhanced by the larger response rate achieved.

At the same time, the sample was redesigned. The goal of the redesign, inspired by Citro and Kalton's (1989) recommendation, was to reduce the number of sampling cells through restratification and to introduce greater homogeneity in sampling rates across cells. The 1991 sample was restratified into 239 cells based on field of degree, sex, and an eight-category demographic group variable defined as: (1) U.S.-born disabled individuals; (2) U.S.-born nondisabled whites; (4) U.S.-born nondisabled Hispanics; (5) U.S.-born, nondisabled Asians; (6) U.S.-born, nondisabled Native Americans; (7) foreign-born U.S. citizens; and (8) foreign-born non-U.S. citizens.

The 1991 sample was initially selected using the 1989 stratification variables and selection probabilities. The need for sample size reduction was discovered *after* the initial selection, and a decision was made to subsample this initial sample after restratification to derive the reduced sample for the 1991 SDR. Thus, the sample design for the 1991 SDR can be considered to be a two-phase sample design. The initial sample became the Phase I sample, which was then restratified resulting into 239 Phase II strata, using a combination of DRF data and SDR questionnaire responses from past surveys. Survey-derived data were used in the Phase II stratification when DRF data were unavailable. This happened most often for the demographic group variable, which used data items not collected in the earlier years of the Survey of Earned Doctorates used to construct the DRF. Thus, the restratification was done using a two-phase sample design. The 1989 strata in the original sample were restratified to create the new 1991 Phase II strata, with the 1989 Phase I strata intersecting the 1991 Phase II strata. A subsampling rate $r_{hh'}$ was used to subsample 1989 sample selections falling in the intersection of old 1989 Phase I stratum h and the new 1991 Phase II stratum h'. These rates were set to equalize, to the greatest extent possible, the ultimate unconditional selection probabilities for the resultant 1991 sample within Phase II strata. For the original Phase I sample, the probability that a unit from stratum h was selected is given by n_h/N_h where n_h and N_h are the sample size and the population size of the hth stratum respectively. The exact probability of initial sample selections made before 1991 is unknown. Restratifications of the SDR sample occurred before 1989 and the sampling probabilities of the units were not recorded. The assumption was made instead that had the true selections probabilities been known, they would have been equal within strata. This assumption cannot be verified. The sampling rate $r_{hh'}$ was calculated as:

$$r_{hh'} = \begin{cases} \dfrac{N_h}{n_h}\dfrac{n_{h'}}{N_{h'}} & \text{if } \dfrac{n_h}{N_h} \geq \dfrac{n_{h'}}{N_{h'}} \\ 1 & \text{if } \dfrac{n_h}{N_h} \leq \dfrac{n_{h'}}{N_{h'}} \end{cases} \tag{1}$$

where n_h, and N_h are the sample size and the population size of the h'th stratum in the Phase II subsample, respectively. Note that no Phase II subsampling was done ($r_{hh'}$=1) when the Phase I sampling occurred at a lower rate than desired for the Phase II subsample.

For the Phase II subsample of the original Phase I sample, the unconditional probability that unit *i* from stratum *h* of the Phase I sample and now classified into Phase II stratum *h'* is in the Phase II subsample is given by:

$$P(\, i \in S_{II}\,) = \frac{n_h}{N_h}\ \ r_{hh'}.$$

From (1) it follows that:

$$P(\, i \in S_{II}\,) = \begin{cases} \dfrac{n_{h'}}{N_{h'}} \text{ if } \dfrac{n_h}{N_h} \geq \dfrac{n_{h'}}{N_{h'}} \\ \\ \dfrac{n_h}{N_h} \quad \text{otherwise} \end{cases}.$$

Thus, unequal selection probabilities occurred within Phase II strata, when the ultimate desired sampling rate from the Phase II stratum was greater than the initial Phase I selection probability for some Phase I strata. Unequal selection probabilities within Phase II strata also resulted from the conversion to integer Phase II subsample sizes prior to sample selection and the later subsampling of mail nonrespondents for follow up by telephone.

4. 1991 Weighting

The basic weight $W_{hi}^{(1)}$ for the *i*th case in the *h*th Phase II stratum was defined as:

$$W_{hi}^{(1)} = \frac{1}{P(\, i \in S_{II}\,)}.$$

Here, $W_{hi}^{(1)}$ represents the basic weight for the mail respondents.

Nonrespondents to the mail survey were subsampled prior to computer-assisted telephone interviewing (CATI) followup. The selection of this mail nonrespondent subsample (about 65 percent of all mail nonrespondents) was done independently of the 1991 sample design. Therefore, the basic weight $W_{hij}^{(2)}$ for the *i*th mail nonrespondent in the *h*th Phase II stratum and the *j*th CATI subgroup can be defined as:

$$W_{hij}^{(2)} = W_{hi}^{(1)} * \ \frac{1}{P_j},$$

where P_j represents the subsampling probability of the *j*th CATI subgroup. Mail nonrespondents who were not subsampled for CATI followup have basic weights of zero.

The next stage was to adjust the basic weight for nonresponse. Nonresponse adjustment cells were created within each Phase II strata using poststratification. A study of the characteristics of mail nonrespondents revealed that they were closer to the characteristics of CATI respondents than mail respondents. Therefore, the decision was made to apply nonresponse adjustments only to the CATI respondents. (In this situation, mail respondents receive a nonresponse adjustment factor of one.) However, there were not enough CATI respondents in some of the nonresponse adjustment cells, and in this situation, all the mail respondents that belonged to the adjustment cell were merged with the CATI respondents. Each nonresponse adjustment cell was defined as one of the *k* subsets formed by poststratification within a given *h*th Phase II strata of the 1991 design. The nonresponse adjustment factor NAF_{hk} for nonresponse adjustment cell *hk* was computed as:

$$NAF_{hk} = \frac{\sum_i n_{hik}^{mr} \; W_{hi}^{(1)} \; + \; \sum_i \sum_j (\; n_{hijk}^{cr} + n_{hijk}^{cnr} \;) W_{hij}^{(2)}}{\sum_i n_{hik}^{mr} \; W_{hi}^{(1)} \; + \; \sum_i \sum_j n_{hijk}^{cr} \; W_{hik}^{(2)}}$$

where n_{hik}^{mr}, n_{hijk}^{cr}, and n_{hijk}^{ncr} are the number of mail respondents transferred to the adjustment cell, the number of CATI respondents, and the number of CATI nonrespondents—all classified by Phase II strata and poststrata—respectively. Poststrata were formed by classifying the Phase II strata by four age groups: (1) less than 35, (2) 35-44, (3) 45-64, and (4) 65 and older.

The initial set of nonresponse adjustment factors were examined and, under these conditions, the cells were collapsed:

- Cells with a nonresponse adjustment factor greater than 2.0.
- Cells with fewer than three respondents and an adjustment factor greater than 1.8.
- Cells with no respondents.

Cells were collapsed until all these rules held true. Collapsing was done across age groups first. When merging within the nonresponse cell did not yield the required adjustment factor or cell size, merges were done across similar degree fields, keeping demographic group and gender the same, or across demographic group, keeping degree field and gender the same. Males and females were never combined. The nonresponse adjustment factors were then calculated for the nonresponse adjustment cells after collapsing.

The final analysis weights were constructed as the product of the basic weight and the nonresponse adjustment factor. The analysis weights were specific to each of the groups of mail respondents and CATI respondents categorized by the Phase II strata, the subsampling group (for the CATI respondents), as well as the nonresponse adjustment groups. Thus, the final weights for mail respondents and CATI respondents are given by:

$$W_{hik}^{(m)} = \; NAF_{hk} \; * \; W_{hi}^{(1)}$$

and

$$W_{hijk}^{(c)} = NAF_{hk} * W_{hij}^{(2)} ,$$

respectively. Note that NAF_{hk} is equal to one for all mail respondents, with the exception of those who were collapsed with CATI respondents when nonresponse adjustment cells were created.

C. The 1993 SDR

The 1993 SDR was also conducted as a mail survey with telephone followup, but for the 1993 SDR all mail nonrespondents were followed. The response rate to the mail survey was about 61 percent; CATI followup boosted response to a final weighted response rate of about 86 percent.

1. 1993 Survey Content

The survey content changed dramatically between 1991 and 1993. The survey instrument was expanded from 8 to 20 pages. The sections on current employment and demographic characteristics were expanded and revised to improve validity and relevance to policy interests. The concept of "employment field " was replaced by that of "occupation." Finally, the reference period was changed from "September" in 1991 to "the week of April 15th" in 1993. Thus, between the 1991 and 1993 surveys, about 20 months had elapsed, as opposed to 31 months between the 1989 and 1991 surveys.

2. 1993 Sampling Frame

For the 1993 SDR, the sampling frame was selected from the DRF using the same three criteria specified for the 1991 SDR.

3. 1993 Sample Design

In 1993, the sample size was increased from the size used in 1991, although it was not up to the 1989 level. The 1993 sample cases were selected from three groups: (1) the 1991 reduced sample; (2) the remaining cases from 1989 who were not in the 1991 reduced sample; and (3) the new 1991-1992 cohort (treated as a separate strata for sampling purposes).

Unlike the 1991 and 1995 surveys, all mail nonrespondents were followed up by CATI in 1993. Therefore, sampled cases had only one probability of selection regardless of whether they responded by mail or CATI.

4. 1993 Weighting

The same nonresponse adjustment cells that were created in 1991 were used in 1993, and the nonresponse adjustment factor NAF_{hk} for each nonresponse adjustment cell *hk*—classified by Phase II strata and poststrata, respectively—was computed as:

$$NAF_{hk} = \frac{\sum_{R} W + \sum_{NR} W}{\sum_{R} W} = \frac{\text{total of basic weights for eligible cases}}{\text{total of basic weights for eligible respondents only}}$$

However, a new set of guidelines was established for collapsing nonresponse adjustment cells to ensure an adequate level of sample representation within each cell. The following cells were collapsed:

- Cells with a response factor greater than 2.0.
- Cells with fewer than 20*(1-*n/N*) respondents, where n_r is the respondent sample size and *N* is the sum of the weights. (Smaller cell sizes were not collapsed if the response rate was 100 percent.)
- Cells with no respondents.

Let $BSCWGT_{hh'}$ denote the basic weight for the mail and CATI respondents in Phase II stratum *h'* and Phase I stratum *h*. Then the final weight for the mail and CATI respondents in cell *hk* is given by

$$FINWGT_{hh'k} = BSCWGT_{hh'} * NAF_{hk} .$$

As in prior years, estimates were produced by summing the final weights of those cases possessing the characteristic of interest.

D. The 1995 SDR

The 1995 SDR was similar to the 1993 SDR, but with some important differences, too. Data collection was by mail, but the 1995 SDR (like the 1991 SDR) followed a subsample rather than all mail nonrespondents. The 1995 SDR also introduced the concept of a "maintenance cut" for the sample design. Prior to 1995, the general approach was to follow all persons selected for the prior survey and to add a sample of new cohorts. Thus, the sample grew over time, except for periodic cuts associated with budget limitations.

1. 1995 Survey Content

The 1995 questionnaire reflected only minor changes made to the question wording of the 1993 instrument and repeated the use of the week of April 15 as the reference period. Two one-time only modules were added to the core questionnaire to obtain information on work history and postdoctoral appointments.

2. 1995 Sampling Frame

The 1995 sampling frame was constructed following the procedures of the 1991 and 1993 SDR, in that it defined the target population as being age 75 or less as opposed to previous surveys which based inclusion on the year the doctoral degree was received.

3. 1995 Sample Design

Prior to 1995, the total size of the SDR sample was not a fixed number. Except for cuts associated with budget constraints, the SDR sample grew in size in each survey cycle because the size of the incoming cohort (new graduates) was typically larger than the size of the outgoing cohort (graduates over age 75). Since 1995, however, the overall sample size for the SDR has been fixed at about 50,000. A "maintenance" cut is taken of the previous cycle's old sample and of the freshly selected sample of recent doctorate recipients to yield a total sample size of 50,000. To implement the maintenance cut, the 1995 SDR sample design divided the target population into three groups:

- The pre-1991 population (degrees received prior to July 1, 1988) represented by SDR selections made prior to the 1991 SDR for whom some required stratification information comes not from the DRF but from previous surveys, yielding a sample design that is two phase with unequal weighting within the second-phase strata (when equal weights are desirable).

- The 1991-1993 population (degrees received from July 1, 1988, to June 30, 1992) represented by SDR selections for the 1991 or later cycles for whom required stratification information is directly available from the DRF.

- The 1995 recent doctorates (degrees received from July 1, 1992 to June 30, 1994) who were sampled directly from the DRF-derived frame, which contained the required stratification information.

These three groups constituted substrata of the desired Phase II strata. Each stratum's sample was allocated to these three substrata in proportion to its population size (a sample-estimated quantity for substratum 1 cases) and the result rounded to the nearest integer. The sample for each substratum was then systematically selected with probability proportional to size (PPS) where the size measure was the unit's basic weight (or a size measure of one for substratum three selections sampled directly from the DRF). Some persons had basic weights so large that they had to be taken into the sample with certainty. This PPS selection tended to make the 1995 sampling probabilities less variable within strata and hence reduced the deleterious effect of unequal weighting associated with 1993 sample selections. In contrast to the 1993 SDR, which followed all mail nonrespondents, the 1995 SDR subsampled mail nonrespondents for CATI followup. This subsampling occurred within Phase II strata and occurred with probability proportional to the basic weight.

4. 1995 Weighting

The weighting plan for the 1995 SDR followed that of the 1993 SDR with one exception. The 1995 SDR subsampling of mail nonrespondents for CATI followup led to the need for two basic

weights: one weight for mail respondents and one weight for subsampled mail nonrespondents. The basic weight was calculated as the inverse of the probability of selection. For the older cohorts, the selection probability was the probability of being selected for the Phase I sample times the probability of being subsampled for the Phase II sample in 1993 times the probability of being subsampled for the initial 1995 mail survey and (for mail nonrespondents) times the probability of being subsampled for CATI followup. As noted earlier, the exact selection probabilities are unknown for selections made prior to 1989. The weighting assumed that the selection probabilities were equal within 1989 strata.

The procedures for nonresponse adjustment followed those of the 1993 SDR.

E. Cross-Survey Similarities and Differences

The above description makes clear that these substantial changes occurred in the SDR in the 1990s:

- The sample design went from a (more or less) stratified, simple random sample in 1989 to a complex two-phase sample in 1991. Additional sample design changes were made in 1995 to improve survey precision and to institute a maintenance cut.

- The 1991 questionnaire was similar in content to the 1989 questionnaire, but with a different reference period and formatting changes that improved item response rates. The questionnaire underwent a major revision in 1993, with only minor wording changes taking place in subsequent years.

- Throughout this period, strata were used to form the weighting classes for nonresponse adjustment. The strata, however, changed dramatically in 1991. In addition, different collapsing rules were adopted in 1993 to produce more stable nonresponse adjustments.

The next section examines the effect of these changes on time-series estimates.

3. WEIGHTING FOR TIME-SERIES ESTIMATION

Given the changes to the Survey of Doctorate Recipients that occurred between 1989 and 1995, these questions arise:

- Are these changes having an adverse affect on time-series analyses?
- If so, is there an estimation approach that can compensate for these differences?

To look into the second issue, we instituted an investigation of reweighting procedures designed to provide greater control over variables that were added as stratification variables in 1991. Normally, the cross-sectional weights constructed for each survey year would be used for time-series analysis. However, design changes may make the 1989 cross-sectional weight not necessarily suited for use in some time-series analysis for these reasons:

- The SDR has historically used strata to define weighting classes in constructing these weights. The addition of new stratification variables in 1991 could have substantial impact on time-series analyses, particularly if the new variables are associated with domains that respond at different rates.
- The dramatic increase in response rates between 1989 and 1991 could have a substantial effect on time-series analyses particularly if mail respondents (the only ones included in 1989) have different characteristics from telephone respondents.

As a consequence, differences between the conventional 1989 and 1991 cross-sectional estimates may reflect true differences confounded with the effects of design changes. For this reason, a special 1989 time-series weight may be needed to remove the confounding effects of the design changes. Given the nature of the sample design and data collection changes, it was only possible to change the 1989 weighting methodology in an attempt to bring the 1989 SDR in greater conformity to the 1991 SDR. In this chapter, we first describe the creation of revised 1989 time-series weights through poststratification and then discuss a further revision of the poststratified weights through benchmarking (see Table 3-1).

A. THE 1989 POSTSTRATIFIED WEIGHT

The first step in the creation of revised 1989 time-series weights was to replicate features of the 1991 sample design in developing weights for the 1989 sample. To do so, we treated the 1987 sample strata as if they were 1989 Phase I strata, and then used the 1991 Phase II stratum definitions to define 1989 Phase II strata. (As in 1991, information from prior SDR surveys and from the DRF was used to classify sample cases into the Phase II strata.) The only difference between the "redefined" 1989 design and the 1991 design was that in 1989 there was no subsampling in Phase II because there was no sample reduction.

Table 3-1. Comparison of conventional and alternate weighting for the 1989 SDR

Weight Component	1989 Original Weights	1989 Poststratified Weights	1989 Benchmarked Weights
Basic weight	Inverse of sampling rate	Inverse of initial sampling rate	Inverse of sampling rate
Nonresponse Adjustment Factor (NAF)	NAF applied to mail respondents within Phase I strata	NAF applied to mail respondents in cells created by crossing the Phase II strata with four age groups: 1. Less than 35 2. 35-44 3. 45-64 4. 65 and up	NAF applied to respondents in cells created by crossing the Phase II strata with four age groups: 1. Less than 35 2. 35-44 3. 45-64 4. 65 and up
Collapsing Rules		Cells were collapsed using the 1991 guidelines: ■ Cells with a NAF > 2.0 ■ Cells with fewer than three respondents and NAF > 1.8 ■ Cells with no respondents	Cells were collapsed using the 1991 guidelines: ■ Cells with NAF > 2.0 ■ Cells with fewer than three respondents and NAF > 1.8 ■ Cells with no respondents
Benchmarking	None	None	Nonresponse-adjusted weight = basic weight * NAF Nonresponse-adjusted weights benchmarked to 1991 control totals by race in five groups: 1. White 2. Black 3. Asian/Pacific Islander 4. Native American 5. Other
Analysis Weight	Basic weight * NAF	Basic weight * NAF	Nonresponse adjusted weight multiplied by the benchmark adjustment factor

Therefore, for the 1989 Phase II sample, the probability that unit *i* from stratum *h* of the Phase I sample, and now classified into Phase II stratum *h'* , is in the Phase II sample S_{II} is given by

$$P(i \in S_{II}) = \frac{n_h}{N_h},$$

which is the probability of selection into the Phase I sample. In other words, this procedure did not change the basic weights for the 1989 sample cases.

The next step was to replicate for 1989 the procedures used in 1991 to create weighting classes. Note that in 1989, poststratification was not used in nonresponse adjustment (adjustment for nonresponse was done using the sampling strata only). For the reweighting, the 1989 Phase II strata were poststratified by four age groups, as the 1991 Phase II strata were, and the same cell collapsing rules were applied.

Next, we created revised 1989 weights by multiplying the basic weights by the nonresponse adjustment factor derived from the poststratification adjustment. These poststratified weights were then used to reestimate selected demographic and employment characteristics. A comparison of these estimates with the original estimates showed race to be most affected by the revision. In terms of proportions, the percentage of white and black doctoral scientists and engineers increased by 0.5 and 0.2 percentage points, respectively. The percentage of Asians dropped by 0.6 percentage points. These changes were relatively large compared with the changes for other variables. The likely explanation for this is that differential nonresponse occurred among these racial groups (blacks typically respond at lower rates than Asians do) which was exacerbated by differences in how the nonresponse adjustment cells were defined in the original 1989 weighting versus the 1989 poststratification weighting which mimicked 1991 weighting classes. In the original 1989 weighting, Asians—who were combined with blacks and other minorities in defining sampling strata—were also combined with blacks and other minorities for nonresponse adjustment. In the revised poststratified weighting, Asians were not combined with other minority groups in defining poststrata.

B. THE 1989 BENCHMARKED WEIGHT

This finding lead to a further refinement in the creation of a revised 1989 weight for potential use in time-series analyses. Because the 1991 survey had a much higher response rate than the 1989 survey (as a result of the 1991 CATI followup and other improvements), the 1991 estimates can be expected to have less nonresponse bias, and hence they should be more accurate. Therefore, the 1991 survey estimates of 1989 racial groups were used to benchmark the 1989 poststratified weights. This was done as follows.

Denote the estimate of the *i*th race/field group in 1989 derived from the 1991 survey as $T_{91}(i)$, and denote the estimate of the *i*th race group in 1989 using the 1989 poststratified weights as $T_{89}(i)$. To derive the benchmarked weights, the 1989 poststratified weights in the *i*th race/field group were adjusted by a factor A_i where:

$$A_i = \frac{T_{91}(i)}{T_{89}(i)}.$$

The difference between the 1991 estimate and the 1989-1990 cohort estimate of the *i*th race group gives the estimate of the *i*th race/field group in 1989 using the 1991 survey. Note that this estimate does not include the out-of-scope cases in 1991 that were in scope for the 1989 survey. An estimate of these cases in 1989 was obtained using the 1989 poststratified weights. Therefore, $T_{91}(i)$ is the sum of the 1991 estimate of the *i*th race group in 1989 and the estimate of 1989 in-scope cases that were out of scope in 1991.

The benchmarked weights changed the percentage distribution by race only slightly from the poststratified estimates (in all groups, less than 0.1 percentage points). However, estimates of population totals changed in more important ways. The total population size was about two percent higher using the poststratified weights as compared with the benchmarked weights. This difference was probably caused by the fact that the 1991 control totals (used for benchmarking) were based on a much higher response rate than that achieved in 1989. This increased response led to the identification of many more out-of-scope sample members, that is, individuals who were living outside the United States or who were deceased. It thereby lowered the size of the estimated in-scope population. Thus, the benchmarked estimates are probably more accurate than the poststratified estimates because they account for the reduction in nonresponse bias that would have been achieved with a higher response rate.

For subsequent analyses, we decided to focus attention only on revised estimates derived using the benchmarked weights.

4. ANALYSIS OF SELECTED VARIABLES

In this chapter, we examine time trends for selected variables from 1987 to 1995. Although most changes occurred during the period 1989-1993, we include survey years 1987 and 1995 to provide a longer trend line for analysis. Tables 4-1 to 4-10 show trend data for selected demographic and employment variables. For 1989, we show two sets of estimates. The first column (labeled "1989 Original") shows totals and percentages estimated using the weights created for cross-sectional estimation in that year. The second column (labeled "1989 Revised") shows totals and percentages estimated using the final set of 1989 revised weights (the "benchmarked" weights).

Variables were chosen for this analysis based on two criteria: (1) they were asked in all five survey years, 1987-1995, and (2) they were asked in a comparable way in those years based on our a priori judgment. This last point merits further discussion. As discussed in Chapter 2, the content of the survey questionnaire and the wording of particular items changed in minor ways during the 1987-1991 period (although the impact of these changes on the survey estimates was sometimes major). In 1993, however, the questionnaire was completely redesigned. The redesign involved changes to the way key questions were asked, and the addition of new question modules on topics of special interest. These changes were made to improve the validity and reliability of the survey data and to improve the relevance of the data for policy making. As a consequence, however, some questions were so dramatically restructured that the possibility that they could be comparable in a time series can be ruled out in advance.[4]

A comparison of 1991 and 1993 questionnaire items is included in Appendix A. Before proceeding with a time-series analysis, analysts are encouraged to consult this comparison to learn how the questions changed over time. Even in the presence of revised weighting to bridge design and response rate differences, wording changes may still preclude time-series analysis. This judgment is left to the individual researcher.

In the time-series tables, we present ten demographic and employment variables that met our criteria for selection. In the remainder of this chapter, we discuss overall trends for these variables, compare 1989 original and revised estimates, and assess the usefulness of the 1989 revised weights in improving the trend lines.

To evaluate the trend lines, we looked for patterns between years that were consistent with the overall trend. For example, the percentage of women scientists and engineers showed a pattern of increase over the 1989-1995 period, so we looked for estimates that were contrary to this pattern, or that showed rates of increase that were inconsistent with other years. Such findings would signal a potential break in the time series that might be caused by the 1991 design changes, the 1989-1991 response rate differences, or the 1993 wording changes. Further, we looked to see if the 1989 revised weights appeared to "smooth" the break, if they made no difference, or if they produced an estimate contrary to the observed trend.

---

[4] These questions and their related analysis variables are government-support status, agencies of support, areas of national interest, energy source, primary energy-related activities, and factors affecting consideration of research abroad (see Appendix A).

Table 4-1. Comparison of weighted counts for degree field of doctoral scientists and engineers, 1987-1995

Degree Field	1995		1993		1991		1989 Revised		1989 Original		1987	
Life Sciences	150,851	27.8%	140,861	27.4%	130,563	27.2%	121,517	27.1%	124,224	27.0%	115,016	26.9%
Engineering	87,004	16.0%	81,560	15.9%	75,094	15.6%	68,566	15.3%	70,228	15.3%	64,481	15.1%
Math/Computer Sciences	31,740	5.9%	29,734	5.8%	26,962	5.6%	24,857	5.5%	25,565	5.6%	23,666	5.5%
Physical Sciences	117,419	21.6%	113,019	22.0%	107,112	22.3%	101,353	22.6%	105,014	22.8%	99,090	23.1%
Social Sciences	155,523	28.7%	148,289	28.9%	140,473	29.3%	131,799	29.4%	135,194	29.4%	125,933	29.4%
Total	542,537	100.0%	513,463	100.0%	480,204	100.0%	448,092	100.0%	460,225	100.0%	428,186	100.0%

Table 4-2. Comparison of weighted counts for gender of doctoral scientists and engineers, 1987-1995

Gender	1995		1993		1991		1989 Revised		1989 Original		1987	
Male	425,930	78.5%	410,191	79.9%	389,674	81.1%	369,404	82.4%	380,069	82.6%	358,805	83.8%
Female	116,606	21.5%	103,272	20.1%	90,530	18.9%	78,688	17.6%	80,156	17.4%	69,381	16.2%
Total	542,536	100.0%	513,463	100.0%	480,204	100.0%	448,092	100.0%	460,225	100.0%	428,186	100.0%

Table 4-3. Comparison of weighted counts for race of doctoral scientists and engineers, 1987-1995

Race	1995		1993		1991		1989 Revised		1989 Original		1987	
White	466,077	85.9%	446,074	86.9%	419,439	87.3%	400,262	89.3%	408,967	88.9%	382,657	89.4%
Black	11,432	2.1%	10,380	2.0%	9,577	2.0%	7,896	1.8%	7,091	1.5%	6,295	1.5%
Asian	62,738	11.6%	54,896	10.7%	47,555	9.9%	37,230	8.3%	41,253	9.0%	36,487	8.5%
Native American	2,195	0.4%	2,112	0.4%	1,785	0.4%	1,534	0.3%	1,727	0.4%	1,525	0.4%
Other	94	0.0%			140	0.0%	160	0.0%	154	0.0%	109	0.0%
No report					1,708	0.4%	1,011	0.2%	1,033	0.2%	1,113	0.3%
Total	542,536	100.0%	513,462	100.0%	480,204	100.0%	448,093	100.0%	460,225	100.0%	428,186	100.0%

Table 4-4. Comparison of weighted counts for Hispanic origin of doctoral scientists and engineers, 1987-1995

Hispanic Origin	1995		1993		1991		1989 Revised		1989 Original		1987	
Hispanic	11,925	2.2%	10,044	2.0%	9,046	1.9%	7,720	1.7%	8,228	1.8%	6,871	1.6%
Not Hispanic	530,511	97.8%	503,419	98.0%	468,961	97.7%	438,089	97.8%	449,679	97.7%	418,243	97.7%
No report					2,197	0.5%	2,283	0.5%	2,318	0.5%	3,072	0.7%
Total	542,436	100.0%	513,463	100.0%	480,204	100.0%	448,092	100.0%	460,225	100.0%	428,186	100.0%

Table 4-5. Comparison of weighted counts for citizenship of doctoral scientists and engineers, 1987-1995

Citizenship	1995		1993		1991		1989 Revised		1989 Original		1987	
U.S. citizens	499,893	92.1%	474,009	92.3%	448,459	93.4%	412,289	92.0%	433,944	94.3%	404,906	94.6%
Foreign citizens	42,643	7.9%	39,454	7.7%	31,632	6.6%	35,719	8.0%	26,228	5.7%	23,241	5.4%
No report					113	0.0%	83	0.0%	53	0.0%	39	0.0%
Total	542,536	100.0%	513,463	100.0%	480,204	100.0%	448,091	100.0%	460,225	100.0%	428,186	100.0%

Table 4-6. Comparison of weighted counts for employment status of doctoral scientists and engineers, 1987-1995

Employment Status	1995		1993		1991		1989 Revised		1989 Original		1987	
Full-time employed	434,251	80.0%	415,994	81.0%	396,118	82.5%	378,730	84.5%	390,021	84.7%	366,668	85.6%
Part-time employed	27,713	5.1%	29,001	5.6%	24,436	5.1%	20,665	4.6%	21,249	4.6%	18,568	4.3%
Postdoctoral appointment	22,814	4.2%	17,871	3.5%	11,240	2.3%	15,653	3.5%	14,611	3.2%	12,166	2.8%
Seeking employment	7,336	1.4%	7,636	1.5%	6,328	1.3%	3,645	0.8%	3,672	0.8%	4,346	1.0%
Retired	40,565	7.5%	34,669	6.8%	34,175	7.1%	24,212	5.4%	25,371	5.5%	20,740	4.8%
Not employed, not seeking	9,857	1.8%	8,291	1.6%	7,601	1.6%	5,150	1.1%	5,261	1.1%	4,761	1.1%
No report					306	0.1%	38	0.0%	40	0.0%	937	0.2%
Total	542,536	100.0%	513,462	100.0%	480,204	100.0%	448,093	100.0%	460,225	100.0%	428,186	100.0%

Table 4-7. Comparison of weighted counts for type of employer of doctoral scientists and engineers, 1987-1995

Type of Employer	1995		1993		1991		1989 Revised		1989 Original		1987	
Educational Institution	234,938	48.5%	221,790	47.9%	202,649	46.9%	212,940	51.3%	216,405	50.8%	205,381	51.7%
Business/self-employed	175,274	36.2%	169,466	36.6%	156,319	36.2%	135,064	32.5%	140,224	32.9%	126,973	32.0%
Non profit*	23,842	4.9%	23,614	5.1%	29,478	6.8%	27,280	6.6%	27,883	6.5%	26,183	6.6%
Government	47,977	9.9%	46,601	10.1%	39,486	9.1%	37,324	9.0%	39,049	9.2%	36,633	9.2%
Other	2,747	0.6%	1,392	0.3%	1,579	0.4%	1,679	0.4%	1,544	0.4%	1,540	0.4%
No report					2,283	0.5%	762	0.2%	776	0.2%	692	0.2%
Total	484,778	100.0%	462,863	100.0%	431,794	100.0%	415,049	100.0%	425,881	100.0%	397,402	100.0%

*All respondents selecting the category "hospital or clinic" from 1987 to 1991 were included in the nonprofit sector in this table. The "hospital or clinic" category was not an option on the 1993 or 1995 questionnaires.

Table 4-8. Comparison of weighted counts for primary work activity of doctoral scientists and engineers, 1987-1995

Work Activity	1995		1993		1991		1989 Revised		1989 Original		1987	
Teaching	106,968	22.1%	101,220	21.9%	96,684	22.4%	101,698	24.5%	103,779	24.4%	102,018	25.7%
R&D*	198,891	41.0%	191,124	41.3%	156,778	36.3%	158,740	38.2%	162,126	38.1%	149,924	37.7%
Management**	57,315	11.8%	63,164	13.6%	67,327	15.6%	66,661	16.1%	69,147	16.2%	64,110	16.1%
Other activities	121,606	25.1%	107,358	23.2%	102,200	23.7%	85,103	20.5%	87,894	20.6%	78,471	19.7%
No report					8,805	2.0%	2,848	0.7%	2,935	0.7%	2,879	0.7%
Total	484,780	100.0%	462,866	100.0%	431,794	100.0%	415,050	100.0%	425,881	100.0%	397,402	100.0%

* Includes applied research, basic research, design, and development.

** Includes management/administration of R&D and management/administration of educational/other programs in 1987, 1989, and 1991; management and administration in 1993; and managing and supervising in 1995.

Table 4-9. Comparison of weighted counts for faculty rank of academically employed doctoral scientists and engineers (excluding postdocs), 1987-1995

Faculty Rank	1995		1993		1991		1989 Revised		1989 Original		1987	
Professor	80,183	37.9%	78,448	38.5%	73,158	38.5%	81,717	41.6%	84,766	42.2%	80,520	41.8%
Associate professor	50,716	24.0%	48,962	24.1%	46,326	24.4%	46,686	23.8%	48,296	24.1%	48,511	25.2%
Assistant professor	42,126	19.9%	41,306	20.3%	36,122	19.0%	38,055	19.4%	37,301	18.6%	35,812	18.6%
Instructor	3,933	1.9%	3,641	1.8%	2,681	1.4%	2,374	1.2%	2,374	1.2%	2,376	1.2%
Lecturer	2,363	1.1%	2,374	1.2%	2,204	1.2%	2,222	1.1%	2,316	1.2%	1,937	1.0%
Adjunct faculty	4,653	2.2%	4,888	2.4%	3,525	1.9%	3,557	1.8%	3,679	1.8%	2,536	1.3%
Other faculty	3,109	1.5%	1,546	0.8%	4,639	2.4%	11,536	5.9%	11,535	5.7%	10,742	5.6%
Not applicable*	24,665	11.6%	22,384	11.0%	11,157	5.9%	4,203	2.1%	4,283	2.1%	4,154	2.2%
No report					9,972	5.3%	6,117	3.1%	6,223	3.1%	5,815	3.0%
Total	211,748	100.0%	203,549	100.0%	189,784	100.0%	196,467	100.0%	200,773	100.0%	192,403	100.0%

*Includes both not applicable at institution and not applicable for position in 1993 and 1995.

Table 4-10. Comparison of weighted counts for tenure status of academically employed doctoral scientists and engineers (excluding postdocs), 1987-1995

Tenure Status	1995		1993		1991		1989 Revised		1989 Original		1987	
Tenured	118,645	56.0%	115,522	56.8%	108,196	57.0%	111,609	56.8%	115,721	57.6%	111,009	57.7%
On tenure track	37,374	17.7%	36,988	18.2%	34,766	18.3%	33,184	16.9%	32,381	16.1%	31,410	16.3%
Not tenured, not on track	18,501	8.7%	17,573	8.6%	15,562	8.2%	18,249	9.3%	18,362	9.1%	15,959	8.3%
Not applicable*	37,229	17.6%	33,465	16.4%	20,862	11.0%	15,252	7.8%	15,591	7.8%	14,665	7.6%
No report					10,398	5.5%	18,171	9.2%	18,718	9.3%	19,360	10.1%
Total	211,749	100.0%	203,548	100.0%	189,784	100.0%	196,465	100.0%	200,773	100.0%	192,403	100.0%

*Includes both not applicable at institution and not applicable for position in 1993 and 1995.

Our evaluations are necessarily subjective because the true population values are unknown. We have only estimates to compare and a decision to make about which appear implausible. We did not consider events in the external environment that might explain inconsistencies because such a discussion would be highly speculative. Further, there are no superior estimates from other data sources that we might use to approximate truth. For this reason, we do not discuss standard errors or the statistical significance of differences between estimates because they do not provide insight into which 1989 estimates are closer to the true value. Rather, we evaluate the trend lines based on the evidence at hand, with the underlying assumption that data in a time series will tend to show smooth patterns of change in the absence of extraordinary outside influences.

A. DEMOGRAPHIC VARIABLES

The demographic variables examined in our analysis are: field of doctoral degree, gender, race, Hispanic indicator, and citizenship.

1. Field of Doctoral Degree

Table 4-1 shows the target population broken out by the field of doctoral degree in the years from 1987 to 1995. Over the eight-year period, there is little change in the percent distribution by field—about 27 percent of the population hold doctorates in the life sciences, 16 percent in engineering, 6 percent in math/computer sciences, 22 percent in the physical sciences, and 29 percent in the social sciences. Further, the 1989 revised weights do not change the distribution in a meaningful way compared with the 1989 original weights. What the 1989 revised weights do change, however, is the size of the total population in that year. The 1989 original population estimate is 460,225 compared with the revised estimate of 448,092. This 2.6 percent difference is probably caused by the fact that the 1991 control totals used for benchmarking the 1989 revised weights were based on a much higher response rate than that attained in 1989. This higher response rate led to the identification of more out-of-scope sample members, that is, individuals who were living outside the United States or who were deceased. Through this identification, the size of the in-scope population was reduced. Thus, the 1989 revised population estimate may be more accurate than the 1989 original estimate because it accounts for the reduction in nonresponse bias that would have been achieved had a higher response rate been attained in 1989.

2. Gender

The overall trend for this variable shows a steadily increasing percentage of women, from 16.2 percent in 1987 to 21.5 percent in 1995. Further, there is little difference between 1989 original and revised estimates. Thus, this variable appears to be stable in time-series analysis, probably because it is less likely to be affected by design or wording changes. Gender was also used for stratification and weighting before and after 1989, and thus was controlled for in the design.

3. Race

The distribution of this variable was perhaps most affected by the 1991-1993 design changes in ways that are not smoothed over using the 1989 revised weights. The overall trend shows small increases in the percentages of blacks in the population and somewhat larger increases in the percentage

of Asians. For Asians, however, the 1989 original estimates appear to be more consistent with that trend than the 1989 revised estimates do. To elaborate, in 1987 the percentage of Asians was 8.5 percent, and there was a steady pattern of increase to 11.6 percent by 1995. However, if 1989 revised weights are used, the percentage of Asians drops slightly from 1987, compared to increasing 0.5 percentage points between 1987 and 1989 when original weights are used. For blacks, the opposite is true. The percentage of blacks using 1989 original weights—1.5 percent—is the same as it was in 1987, but the percentage derived using the 1989 revised weights increases to 1.8 percent—more in keeping with the 2.0 estimate in 1991. This correspondence is to be expected, however, since 1991 control totals by race were used to benchmark 1989 estimates when creating the revised weights. Thus, the 1989 revised weights appear to work better in producing a smooth trend for blacks than for Asians.

The above observation may be because the 1989 revised weights treated blacks and Asians separately in weighting for nonresponse (in calculating the 1989 original weights the two groups were combined.) Because Asians respond at higher rates than blacks, separating the two meant that Asians carried lower weights in the revised estimates than in the original estimates. Even with benchmarking to 1991 control totals by race, the reduction in weights for Asians was large enough to result in a small decline in the estimated percentage of Asians in the population. This decline is contrary to the overall trend.

4. Hispanic Indicator

The distribution of this variable shows a small but steady progression in the percentage of Hispanic doctoral scientists and engineers, from 1.6 percent in 1987 to 2.2 percent in 1995. This variable is similar to gender in that there is little difference between 1989 original and revised estimates. Overall, the trend appears to be smooth regardless of the 1989 weight used.

5. Citizenship

Citizenship is another variable for which the 1989 revised weights produce inconsistent results. The overall trend shows an increasing percentage of foreign citizens in the population, from 5.4 percent in 1987 to 7.9 percent in 1995. Compared with 1987 (5.4 percent), the estimated percentage of foreign citizens is 5.7 percent in 1989 (original) and 6.6 percent in 1991. Using the 1989 revised weights, however, the 1989 percentage jumps to 8.0 percent. That increase may be related to better control of nonresponse bias in the revised estimates (foreign citizens typically respond at lower rates than do U.S. citizens), but it is difficult to know why that percentage then drops to 6.6 percent in 1991 when response rates were similarly high.

B. EMPLOYMENT VARIABLES

The employment variables examined in our analysis are: employment status, type of employer, primary work activity, faculty rank, and tenure status.

1. Employment Status

Employment status shows inconsistencies in the time series that are not related to the difference between 1989 original and revised estimates. Indeed, the 1989 original and revised estimates are very similar in terms of percentages. Rather, a break occurred between 1987-1989 and 1991-1995, regardless of which 1989 estimates are used. To illustrate, in 1989 the estimated percentage of those unemployed but seeking employment is 0.8 percent. This percentage increases to 1.3 percent in 1991 and remains fairly stable through 1995. There is an even larger increase in the percentage of retired individuals—from 5.4 percent in 1989 to about 7.1 percent or more in 1991 and beyond. These increases are comparatively large and are not easily explained by labor market events. Further, no changes in question wording occurred between 1989 and 1991 that might account for these differences.

There is also an inconsistent trend in the number reporting postdoctoral appointments—ranging from 2.3 percent in 1991 to 4.2 percent in 1995. The method used to identify postdocs changed considerably over these years, however (see Appendix A), and we do not recommend using these data to draw conclusions about the labor market for postdocs.

2. Type of Employer

The overall trend for employer type shows decreasing percentages employed by educational institutions—from 51.7 percent in 1987 to 48.5 percent in 1995—and increasing percentages employed by businesses or self-employed—32.0 percent in 1987 compared with 36.2 percent in 1995. There is little difference between 1989 original and revised estimates, and the trend is smooth for most employer types.

3. Primary Work Activity

The categories used to define primary work activity have been modified frequently over the years so that collapsing is necessary for time-series analysis. Generally, the percentage reporting teaching declined from 1987 to 1995, correlating with the decline in the percentage who were academically employed. The percentage reporting research and development as a primary work activity generally increased, although in a relatively large way between 1991 and 1993 (36.3 percent and 41.3 percent, respectively). This increase is probably related to the question redesign in 1993 and may preclude the use of this variable in time-series analysis including pre- and post-1993 data.

4. Faculty Rank

The most obvious inconsistencies in the distribution of this variable occur in the "other faculty," "not applicable," and "no report" categories and are probably related to changes in question wording. For "other faculty," the percentage dropped from 5.9 percent in 1989 (revised) to 2.4 percent in 1991, and further still after that. On the other hand, the percentage in the "not applicable" category increased from 2.1 percent in 1989 (revised) to 5.9 percent in 1991, and then to 11.0 percent in 1993. This increase probably occurred between 1991 and 1993 because in 1993 the "not applicable" category was divided into two categories: "not applicable for my position" and "not applicable at this institution." This division may have caused more respondents to select one of these options.

The differences between 1989 and 1991, however, are not so readily explained. As noted earlier, the percentage responding as "other faculty" decreased by over three percentage points between 1989 and 1991, and the percentages in the "not applicable" and "no report" categories increased by two percentage points or more. Because there were no wording changes between 1989 and 1991, these comparatively large changes occurred for unknown reasons.

5. Tenure Status

A similar break occurred in the distribution of the tenure status variable between 1989 and 1991, and in the same categories. For example, the percentage reporting that tenure status was "not applicable" increased from 7.8 percent in 1989 to 11.0 percent in 1991, when the category was split into two categories. In contrast to faculty rank, however, the percentage of "no reports" decreased—from 9.2 percent in 1989 (revised) to 5.5 percent in 1991. This may be related to the improved formatting of the question in 1991 that caused fewer respondents to accidently skip the tenure status question, a follow-up question to faculty rank.

For the other response categories—tenured; on tenure track; not tenured, not on track—the time trend is fairly smooth. Thus, most of the differences between 1989 and 1991 occurred between the "not applicable" and the "no report" categories. These categories both total around 17 percent in each year, despite their dissimilar distributions.

5. CONCLUDING REMARKS

This report discusses three critical influences on the comparability of data from the Survey of Doctorate Recipients in time-series analysis: (1) the 1991 design changes, (2) the pre- and post-1989 response rate differences, and (3) the 1993 questionnaire redesign. Given these changes, it can be seen that using the SDR in time-series analysis spanning the years from 1987 to 1995 is problematic. At the start of our analysis, we identified ten variables that might be suitable candidates for time-series analyses. After looking at the trend lines, we found that only four showed consistent patterns over time: degree field, gender, Hispanic origin, and type of employer. The first three are demographic variables used to stratify the population for sampling and weighting adjustments; therefore, variability is better controlled than it is for type of employer. Nonetheless, given the inconsistencies in the other variables, it is difficult to determine whether the consistent patterns observed in these variables are indicative of reality or merely fortuitous.

The 1989 revised weights were created in an attempt to bridge the design differences between 1989 and 1991 and beyond. Further, they were created to address the potentially serious problem of nonresponse bias in 1989 estimates that were based on a response rate of only 55 percent.[5] The revised weights do not appear to have fully met these objectives. For some variables, the 1989 revised estimates vary little from the original estimates, at least in terms of percentages. For others, the 1989 revised weights change the distributions in ways that improve comparability with 1991 at the expense of 1987, or visa versa. Further, within variables the revised weights sometimes appear to improve consistency in some categories but to worsen consistency in other categories (see, for example, race). For all variables, the 1989 revised weights significantly reduce the size of the doctoral population in that year compared with 1989 original weights. Moreover, estimates based on 1989 revised weights will not agree with previously published estimates for that year and analysts will have to be prepared to explain the differences.

Particularly disturbing to us was the lack of a consistent pattern. We expected that the 1989 original weight would produce a more consistent estimate of change from 1987 to 1989, while the 1989 revised weight would produce more consistent estimates of change from 1989 to 1991. While this was usually true, it was not consistently so, particularly for 1989 to 1991 comparisons. The 1987 and 1989 estimates are based on comparable sample designs, questionnaires, and data collection procedures. The 1989 original weights reflect these similarities. In contrast, the 1991 design is quite different from the 1989 design, but the revised weights attempt to compensate for the differences. The mixed results of our analysis suggest that the 32 percentage point difference in the response rates between 1989 and 1991 may indicate nonresponse bias in the 1989 SDR that is too complex to be removed by the revised weighting procedures.

Another point to consider is that the variables in this analysis were frequently aggregated to a high degree. Primary work activity, for example, was collapsed into four categories from over 12 that appear on the original questionnaires. This collapsing was done to control for changes in the number and type of response categories that were listed in a given year, while preserving those categories that remained consistent. The category "other activities," for example, includes a different combination of

[5] At the start of this project, we considered developing a methodology that might be used to correct nonresponse biases in years prior to 1989 as well. It was clear as the project progressed, however, that this was neither practical nor useful.

categories in most years, while the category "teaching" remains the same across years. If analysts disaggregate these variables for more detailed analysis, then further inconsistencies will emerge.

Finally, weighting adjustments cannot control for all the differences between survey years. The redesign of the questionnaire in 1993, and more minor wording changes in predecessor years, most likely did affect response distributions. Variables such as employer type and employment status are based on questions that were reworded in 1993. The effect of these wording changes is unknown.

On the positive side, there is comparability on subsets of survey years. For example, there were comparatively few design or wording changes between 1987 and 1989 or between 1993 and 1995 Further, the 1993 questionnaire changes to core items will likely remain intact through the remainder of this decade. This means that data from 1993 onward should be comparatively well-suited for time-series analysis.

Including 1989 in a time series, however, is discouraged. When 1989 must be included, the analyst is advised to use 1989 revised weights if the comparisons are forward in time and 1989 original weights if the comparisons are backward in time. Although this strategy does not bridge content differences or solve all anomalies, it tends to be an improvement over ignoring the design and response rate differences.

As a first step in any time-series analysis, analysts should review this document and make decisions about what variables can be compared and in what years. Another factor to consider is the intended use of the data, and the consequences of the decision-making the data will inform. In publishing time-series estimates spanning 1989-1995, a methodological note explaining the limitations of the estimates is called for, as well as acknowledgment that estimates based on 1989 revised weights may differ from those published by other sources.

APPENDIX A

A COMPARISON OF 1991 AND 1993 QUESTIONNAIRE ITEMS

Prepared by

Prudence Brown
National Research Council
1998

Table 1. Comparison of 1993 and 1991 SDR questionnaire items

93 item #	Description	91 item #	Data file variable/comments
A1	working for pay or profit	1	Employment status, see NOTE 1
A2	looking for work	1	Employment status, see NOTE 1
A3	reasons not working	na	
A4	date last worked	na	
A5	last occupation	na	
A6	last occupation code	na	
A7	full-time/part-time	1	Employment status, see NOTE 1
A8	reasons part-time	3	Not comparable—list of reasons different, 91 is mark one, 93 is mark all that apply
A9	previously retired	na	
A10	principal employer name/location	6	
A11	postdoc status	1	Postdoc status, see NOTE 2
A12	educational institution indicator	7	
A13	type of educational institution	7	Type of employer, see NOTE 3
A14	faculty rank	11a	Faculty rank, see NOTE 4
A15	tenure status	11b	Tenure status, see NOTE 5
A16	type of noneducational institution	7	Type of employer, see NOTE 3
A17	occupation of principal job	na	
A18	occupation code of principal job	na	
A19	manager indicator	na	
A20	expertise in science	na	
A21	license	na	
A22	relationship of degree to job	na	

Table 1. Comparison of 1993 and 1991 SDR questionanire items (continued)

93 item #	Description	91 item #	Data file variable/comments
A23	reasons working out of field	na	
A24	most important reason out of field	na	
A25	work activities	na	
A26	primary and secondary activity	12	Primary/secondary work activity, see NOTE 6
A27	supervisory duties	na	
A28	number supervised directly and indirectly	na	
A29	annual salary	14	Annual salary, see NOTE 7
A30	salary based on full-time	na	
A31	government support of work	17a	Not comparable—91 time frame is past year, 93 is reference week
A32	agencies of support	17b	Not comparable—91 time frame is past year, 93 is reference week
A33	areas of national interest	19	Not comparable—categories are different
A34	energy source	22	Not comparable—categories are different
A35	primary energy-related activity	24	Not comparable—categories are different
A36	second job status	na	
A37	occupation of second job	na	
A38	occupation code of second job	na	
A39	salary for second job	na	
A40	relationship of second job to degree	na	
B1	working in 1988	na	
B2	same employer in 1988	na	
B3	principal employer in 1988	na	
B4	educational institution in 1988	na	
B5	noneducational institution in 1988	na	
B6	same occupation in 1988	na	
B7	occupation in 1988	na	
B8	occupation code in 1988	na	

Table 1. Comparison of 1993 and 1991 SDR questionnaire items (continued)

93 item #	Description	91 item #	Data file variable/comments
B9	same employer and same occupation	na	
B10	reasons for changing employer or occupation	na	
B11	research abroad	18a	
B12	consider research abroad	na	
B13	factors affecting consideration of research abroad	18b	Not comparable—91 asks for primary and secondary reason, 93 asks to rate all reasons great deal/somewhat/not at all
C1	years of professional work experience	16	Not comparable—93 asks for years of experience full-time and part-time with no time frame, 91 asks for FTE years since Ph.D.
C2	professional meetings	na	
C3	professional society membership	na	
C4	work-related workshops	na	
C5	areas of work-related workshops	na	
C6	reasons for training activities	na	
C7	most important reason for training activities	na	
D1	highest degree since doctorate	na	
D2	field of study for degree	na	
D3	year of degree	na	
D4	school-related costs for degree paid by employer	na	
D5	college courses taken	na	
D6	reasons for taking courses	na	
D7	field of study for courses	na	
D8	school-related costs for courses paid by employer	na	
E1	birth date	na	
E2	birth place	na	
E3	rural community	na	
E4	parent's education level	na	

Table 1. Comparison of 1993 and 1991 SDR questionnaire items (continued)

93 item #	Description	91 item #	Data file variable/comments
E5	Hispanic indicator	27a	frame variable
E6	Hispanic type	27b	frame variable
E7	race	26	frame variable*
E8	gender	na	
E9	citizenship	25a	
E10	country of citizenship	25b	
E11	in U.S. to stay	na	
E12	living in U.S.	na	
E13	marital status	28	Marital status, see NOTE 8
E14	employment status of spouse	na	
E15	requirements of spouse's job	na	
E16	children	na	
E17	number of children	29	91 categories are under 6, between 6 and 17; 93 categories are under 6, 6-11, 12-17, 18 or older
E18	disabilities	30-33	
E19	age at onset of disability	na	

*Race was asked on the 1993 SDR questionnaire in order to fill in missing data.

Table 2. Comparison of weighted counts for type of employer of scientists and engineers, 1991 and 1993

1993 cagetories	Number	Percent	1991 categories	Number	Percent
2-Year college/junior college/ tech inst	6,627	1.4%	Junior college/2-year/tech inst	6,211	1.5%
4-year college/university	155,300	33.6%			
Health related schools	1,728	0.4%			
University-affiliated research institute	19,432	4.2%			
			4-year college	29,828	7.0%
			University	132,053	30.9%
Medical school	33,611	7.3%	Medical school	29,344	6.9%
Elem/middle/sec school	5,008	1.1%	Elem/middle/sec school	4,254	1.0%
Other educational institution	84	0.0%			
Private for-profit company	128,370	27.7%			
Self-employed, incorporated	12,822	2.8%			
			Business/industry	117,984	27.6%
Self-employed, not incorporated	28,274	6.1%	Self-employed	37,484	8.8%
Private not-for-profit organization	23,614	5.1%			
			Private foundation	2,003	0.5%
			Other nonprofit organization	13,599	3.2%
Local government	4,217	0.9%	Local government	2,966	0.7%
State government	8,588	1.9%	State government	6,897	1.6%
U.S. military	2,127	0.5%	U.S. military	2,114	0.5%
U.S. government, civilian	31,669	6.8%	U.S. government, civilian	27,354	6.4%
Other employer type	1,392	0.3%			
			Hospital or clinic	13,822	3.2%
			Foreign government	43	0.0%
			Other employer type	1,536	0.4%
Total	462,863			427,492	

Table 3. Comparison of weighted counts for primary work activity of scientists and engineers, 1991 and 1993

1993 Cagetories	Number	Percent	1991 Categories	Number	Percent
Teaching	101,220	21.9%	Teaching	96,147	22.4%
Applied research	92,958	20.1%	Applied research	71,373	16.6%
Basic research	64,455	13.9%	Basic research	60,448	14.1%
Development	23,271	5.0%	Development	18,236	4.3%
Design	10,439	2.3%	Design	6,337	1.5%
Management/administration	63,164	13.6%			
			Management of R&D	33,197	7.7%
			Management of ed prog/other	33,983	7.9%
Computer applications	18,684	4.0%	Computer applications	5,867	1.4%
Professional services	57,611	12.4%	Prof. services to individuals	39,542	9.2%
Production/operations/maint	2,200	0.5%	Operations-produc/maint/constr	3,622	0.8%
Sales/purchasing/marketing	6,512	1.4%	Sales/mktg/purch/cust relat	6,774	1.6%
Quality or productivity mgmt	4,096	0.9%	Quality control/testing/eval	4,585	1.1%
Acounting/finance	4,082	0.9%			
Employee relations	3,731	0.8%			
Other activities	10,441	2.3%			
			Consulting	18,515	4.3%
			Report/technical wrtg &edit	6,425	1.5%
			Statistical work	3,343	0.8%
			Multiple responses	195	0.0%
			Other activities	12,561	2.9%
			No report	7,866	1.8%
Total	462,864		Total	429,016	