Contract No: 9313060 and SRS-9531746
MPR Subcontract No.: NSCF-8014-98-001
MPR Reference No.: 8491-100

CURRENT AND ALTERNATIVE DESIGNS FOR THE SURVEY OF DOCTORATE RECIPIENTS

TASK 1 FINAL REPORT

June 30, 1998

Authors:

Brenda G. Cox
Susan B. Mitchell
Ramal Moonesinghe

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
MPR 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

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

List of Tables

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, federal agencies, universities, 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 U.S. system of graduate education has changed dramatically in the 25 years that the survey has been conducted. New multi-disciplinary programs have been established while some traditional fields of study have declined in enrollment. Many more doctorate recipients today are female or foreign than in the 1970s. Employment patterns have also undergone striking changes. Increasingly, doctoral scientists and engineers are accepting positions in business and industry, as opposed to more traditional academic appointments. In the 1970s, the pressing issue was whether the nation's supply of doctoral scientists and engineers was keeping pace with the increasing demands for these highly trained professionals. Today, there are perceptions of an oversupply in some fields and the use of serial postdoctoral appointments as substitutes for permanent positions. The SDR has provided an information base that allows policymakers to assess the relative importance of these issues and concerns and to determine what courses of action should be considered.

As the policy interests underlying the SDR have evolved over time, so has its survey sampling and estimation procedures. In response to changing policy needs and project funding as well as advances in survey methods, the sample design has been modified at almost every two-year iteration over the last 25 years. Today, the SDR design is a patchwork product with unequal weights within strata, probabilities of selection known only approximately, and an allocation that may be far from optimal. Due to its longitudinal nature, sample members, once selected, are contacted for interview every two years until they reach the age of 76. The level of burden imposed on cooperating individuals is large; some individuals will be asked to complete 25 or more interviews before they die or become age ineligible. Increasing levels of nonresponse and the potential for other nonsampling problems are a concern for the SDR under its present design.

In this report, we describe the current SDR design, its design strengths and weaknesses, and outline potential strategies for its redesign.[1] Two alternative designs are presented for the SDR: (1) a "refreshed" panel design and (2) a rotating panel design. The refreshed panel design offers a method for updating the current sample on a one-time only basis to be phased in over a period of time, say ten years. The rotating panel design offers a method for replacing a portion of the survey sample every two years. Both methods can incorporate other design changes, such as more efficient sample allocation and less unequal weighting. The rotating design has somewhat more flexibility to respond to changing policy needs in the future.

---

[1] This report deals only with the portion of the SDR that provides data on U.S.-earned doctorates in science and engineering. Through 1989, SDR included doctorates who were employed in an S&E field or had a foreign-earned doctorate in S&E. Through 1995, SDR also included U.S.-earned doctorates in the humanities. These non-S&E doctorates were sampled in separate strata and their exclusion has no impact on the findings of this investigation.

To aid in understanding the present SDR design, we provide a historical background for the survey, including its goals at inception and how the sampling and survey methodologies have changed over the survey's 25-year history. Then, we describe the features of the 1995 SDR, the latest year for which complete survey outcomes and population estimates are presently available. Next, we present the design features of the refreshed and rotating panel designs (respectively), including a discussion of each design's strengths and weaknesses. Finally, we summarize what NSF needs to consider in evaluating the current design approach vs. the two alternative designs for the SDR.

2. HISTORICAL BACKGROUND

The Survey of Doctorate Recipients (SDR) collects data on the supply and utilization of individuals who are trained in the United States to the doctoral level in science and engineering. Descriptive estimates of this population's employment and demographic characteristics are widely used to address education and human resource issues by policy makers at NSF, in academe, and in industry. The survey data provide a basis for evaluating the stock of highly trained personnel in the United States, for understanding the nation's educational needs, and for measuring the nation's capacity for research and development. Doctoral scientists and engineers play vital roles in the nation's economy by conducting basic research, performing R&D in such areas as health and national defense, and training the nation's future scientists and engineers (National Science Board 1996).

Since 1973, the SDR has surveyed doctoral scientists and engineers every two years. The SDR collects information on employment status, employment sector, primary work activity, annual salary, and academic rank and tenure status, to name a few of the most widely used variables. The survey is longitudinal; that is, the individual members of the survey panel are resurveyed every two years. This design affords the opportunity to study the career patterns of participants, including mobility between occupations, employment fields, and sectors. The data permit not only comparative analysis across science and engineering fields but also across time.

A. Origins of the SDR

The act establishing NSF in 1950 directed the agency to maintain "a register of scientific and technical personnel and in other ways to provide a central clearinghouse for information covering all scientific and technical personnel in the United States, including its territories and possessions" (NSF Web Site 1998). NSF responded by creating the National Register of Scientific and Technical Personnel, which contained information about the location, fields of specialization, and work activities of scientists and engineers trained to different degree levels. The purpose of the register was, in part, to monitor the nation's supply of scientific and engineering personnel in preparation for a "national emergency."

The National Register was discontinued in 1970, but the need for information about the supply and characteristics of doctoral personnel remained. Working with the National Research Council, the NSF established the SDR in 1973 as a means for obtaining current employment data on doctoral-level personnel. This population was defined as individuals living in the United States who held doctoral degrees in science or engineering fields (whether earned in the United States or not), and individuals who held doctorates in other fields but were employed in science and engineering. This population—to become the SDR's sampling frame—was identified from two main sources:

- *Doctorate Records File (DRF).* The DRF is an ongoing census of all research doctorates earned in the United States since 1920. From 1958 on, information for the DRF has been collected through the annual Survey of Earned Doctorates (SED), a questionnaire distributed to doctoral candidates at the time they complete their degree requirements. The questionnaire collects data on field of degree, gender, citizenship status, and race/ethnicity. Coverage of the population of U.S.-earned doctorates is thought to be high, around 98 percent, and the response rate is typically 95 percent or better. For those who do not respond, partial records are compiled from information found in commencement bulletins and lists furnished by the graduate schools. The

DRF is used to identify the universe of doctorate holders who earned their degrees in the United States.

- *The National Register of Scientific and Technical Personnel.* The National Register was used primarily to identify two other groups for the SDR sampling frame: (1) individuals who obtained doctoral degrees in nonscience fields but were employed as scientists and engineers (known as "field-switchers") and (2) individuals who earned science and engineering doctorates in foreign countries but were working in the United States (known as "foreign-earned doctorates"). Coverage of these two groups, however, was incomplete. Only two-thirds of the number of foreign Ph.D.s working in the U.S. were estimated to have been identified, and coverage of the field switchers could not be determined (National Research Council 1972).

Because of this coverage problem, several other sources were used to identify eligible population members: *American Men of Science*, U.S. college catalogs, and the *National Faculty Directory*. In addition, the 25 largest U.S. corporations were asked to provide the names and addresses of employees with foreign-earned doctorates, but response was varied.

Because the target population was defined as those of employable age, a further selection criterion was applied to individuals identified through this search. A 42-year time interval from the time of degree award to the survey year was set, and those who had earned their degrees between January 1, 1930 and June 30, 1972 were designated for inclusion in the frame. The final frame consisted of 261,393 doctoral scientists and engineers who were either educated or working in S&E in the United States.

The sampling strategy for the SDR was stratified random sampling, which offered improved precision compared with simple random sampling. The frame was stratified according to year of doctorate, field of degree, sex, size of doctoral institution, and segment (comprising three classes: (1) science and engineering doctorates from U.S. institutions, (2) field switchers, and (3) foreign-earned doctorates). The stratification variables were selected to conform with major reporting domains and to control variance associated with predictive variables. To improve the reliability of estimates of small subpopulations (for example, women engineers) variable sampling rates were applied across strata. Consequently, some small cells were sampled at 100 percent, while others were sampled at 18 percent. A total sample of 56,096 was selected.

Because the survey was to be conducted by mail, the next step was to locate addresses for the sample cases. Primary sources of address information were the DRF, the National Register, the *National Faculty Directory,* alumni offices, departments of baccalaureate and doctoral institutions, and professional societies. Of the sample, addresses were found for 94 percent.

The 1973 survey questionnaire was mailed to sample members in March, and three follow-up mailings to nonrespondents occurred in May, August, and October. The questionnaire consisted of 18 questions (2 pages) covering such topics as employment status, employment sector, academic rank, annual salary, and a number of demographic questions. The response rate was about 75 percent.[2]

---

[2] This response rate is the number of eligible respondents divided by the estimated number of eligible sample cases.

B. Changes Since 1973

The SDR has undergone a number of changes since its inception, despite being a longitudinal survey where consistency is critical for making comparisons across time.[3] These changes resulted from four primary sources: (1) changing policy interests over the 25-year life span of the survey, (2) changes in the survey environment, (3) advances in survey methodology, and (4) changes in project funding levels. For example, when the SDR began, the primary policy concern was whether the United States had enough doctorates to maintain its economic competitiveness. By 1993, the focus had shifted to whether the nation was producing too many doctorates, and whether this oversupply was leading to unemployment. A mail survey in 1973 could achieve a 75 percent response rate without telephone follow-up; by 1995, such response rates were no longer possible, at least 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 a tension between maintaining data comparability and improving data quality and survey operations.

The above discussion provides a context for understanding the origin of the changes to the survey between 1973 and 1995. These changes, displayed in Tables 1 and 2, were made to the target population, stratification variables, population size, and sample size. In addition, there were changes to the data collection method, questionnaire content and format, survey reference date, and the response rate achieved. Following are highlights of the design:

1. Target Population

- After 1973, the sampling frame of foreign-earned doctorates and field-switchers was not updated because of difficulty in identifying these individuals. Coverage of these individuals became increasingly incomplete as older cohorts were dropped from the frame and newer cohorts were added.

- In 1977, two changes occurred: (1) humanities doctorates were added to the frame and the Survey of Humanities Doctorates was initiated, and (2) foreign students who graduated from U.S. institutions but whose planned postdoctoral location was foreign (learned through the SED) were excluded from the frame. This exclusion was done because of the perceived high probability these individuals would not become part of the U.S. labor force, the population of interest.

- In 1987, the frame was expanded to include individuals who had graduated within 44 years of the survey, instead of 42 years. This expansion, made to increase the number of cases available for studying retirement patterns and trends, was continued in 1989.

---

[3] This situation is not unusual for surveys done over such a long time period. To some degree, the changes derive from the need to improve data collection to achieve current information needs, which cannot be met as effectively with a static longitudinal design.

Table 2-1. Changes to the target population, stratification variables, population size, sample size, and nonrespondent sample size

Year	Target Population	Stratification Variables	Population Size	Sample Size	Mail Nonrespondent* Follow-up Rate
1973	Doctorates who earned degrees between January 1, 1930 and June 30, 1972 and belonged to one of the following segments: 1. science/engineering doctorates from U.S. institutions 2. nonscience/nonengineering doctorates from U.S. institutions working as scientists and engineers 3. science/engineering doctorates from foreign institutions.	• field of degree or field of employment when degree unavailable • cohort (year of degree) • size of doctoral institution (as measured by the number of doctoral degrees) • gender • segment	261,393	56,096	NA
1975	Scientists and engineers who earned doctoral degrees between January 1, 1932 and June 30, 1974. New additions made only to segment 1 after 1973 (segments 2 and 3 not updated after 1973).	• minority status added (known for post-1973 doctorates) • size of doctoral institution dropped	295,970	62,471	NA
1977	Scientists and engineers who earned doctoral degrees between January 1, 1934 to June 30, 1976. Doctorates in humanities from U.S. institutions who earned degrees between January 1, 1930 and June 30, 1976 were added to frame. Foreign citizens, who graduated from U.S. institutions but whose planned postdoctoral location was foreign, excluded from sampling frame after 1975.		Science/ Engineering 328,351	Science/ Engineering 68,532	0%

Table 2-1. Changes to the target population, stratification variables, population size, sample size, and nonrespondent sample size (continued)

Year	Target Population	Stratification Variables	Population Size	Sample Size	Mail Nonrespondent* Follow-up Rate
1979	Scientists, engineers, and humanists who earned doctoral degrees between January 1, 1936 and June 30, 1978.	• citizenship status added (known for post-1957 doctorates) • minority status collapsed from white, Asian, other, and unknown to white/unknown and minority • the number of field strata was reduced	Science/ Engineering 359,041	Science/ Engineering 41,763	0%
1981	Scientists, engineers, and humanists who earned doctoral degrees between January 1, 1938 and June 30, 1980.		Science/ Engineering 389,671	Science/ Engineering 50,019	0%
1983	Scientists, engineers, and humanists who earned doctoral degrees between January 1, 1940 and June 30, 1982. Doctorates who earned degrees in audiology and speech pathology added to sampling frame in 1983.		Science/ Engineering 423,469	Science/ Engineering 54,886	0%
1985	Scientists, engineers, and humanists who earned doctoral degrees between January 1, 1942 and June 30, 1984.		Science/ Engineering 455,388	Science/ Engineering 59,253	0%
1987	Scientists, engineers, and humanists who earned doctoral degrees between January 1, 1942 and June 30, 1986. 42-year time limit expanded to 44 years		Science/ Engineering 492,072	Science/ Engineering 65,867	0%
1989	Scientists, engineers, and humanists who earned doctoral degrees between January 1, 1942 and June 30, 1988.		Science/ Engineering 532,132	Science/ Engineering 73,611	0%

Table 2-1. Changes to the target population, stratification variables, population size, sample size, and nonrespondent sample size (continued)

Year	Target Population	Stratification Variables	Population Size	Sample Size	Mail Nonrespondent* Follow-up Rate
1991	Scientists, engineers, and humanists who are under 76 years of age, and who earned degrees between January 1, 1942 and June 30, 1988. Science/engineering doctorates who earned degrees from foreign institutions and nonscience/nonengineering doctorates working as scientists or engineers dropped from sampling frame after 1989.	redesign of stratification variables into • field of degree • gender • category "group" variable 1. U.S. born-disabled 2. U.S. born-white 3. U.S. born-black 4. U.S. born-Hispanic 5. U.S. born-Asian 6. U.S. born-Native American 7. Foreign-born U.S. citizen 8. Foreign-born non-U.S. citizen	Science/ Engineering 530,215	Science/ Engineering 37,996	60%
1993	Scientists, engineers, and humanists who are under 76 years of age, and who earned degrees between January 1, 1942 and June 30, 1988.		558,726	49,228	100%
1995	Scientists, engineers, and humanists who are under 76 years of age, and who earned degrees between January 1, 1942 and June 30, 1988.		594,275	49,829 maintenance cut introduced	60%

* The percentage of nonrespondents to the mail survey who were subsampled for CATI followup.

Table 2-2. Changes to the data collection method, content and format, reference date, and response rate

Year	Data Collection Method	Content/Format	Reference Date	Response Rate
1973	Four waves of a mail survey	18 questions; 3 pages; designed for "mark sensing"	Date received in 1973 and March 31, 1972	74.6% *
1975	Three waves of a mail survey	20 questions; 4 pages; designed for manual coding; major content changes	Week of February 9-15, 1975 and week of February 10-16, 1974	70.3%
1977	Two waves of a mail survey	17 questions; 4 pages; designed for direct key entry; major content changes; modified for administration to humanities sample	Week of February 6-12, 1977.	63.3%
1979	Four waves of a mail survey	24 questions; 4 pages; major content changes	Week of February 11, 1979	65.7%
1981	Three waves of a mail survey; final wave used shortened questionnaire	25 questions, 4 pages; major content changes	February 1981	63.4%
1983	Four waves of a mail survey; third wave used shortened questionnaire; fourth wave used postcard questionnaire of key items	20 questions; 4 pages; minor content changes	February 1993	66.0%
1985	Three waves of a mail survey; final wave used shortened questionnaire	27 questions; 4 pages	February 1985	63.4%
1987	Three waves of a mail survey; second wave used shortened questionnaires (two lengths) as part of experiment to test the effect of questionnaire length on response	27 questions; 6 pages	February 1987	58.2%
1989	Three waves of a mail survey; no shortened questionnaires were used	27 questions; 6 pages; minor content changes	February 1989	55.2%
1991	Three waves of a mail survey followed by telephone interviewing (CATI) of nonrespondents	35 questions; 8 pages; minor content changes but major changes to format	September 1991	87% **

Table 2-2. Changes to the data collection method, content and format, reference date, and response rate (continued)

Year	Data Collection Method	Content/Format	Reference Date	Response Rate
1993	Two waves of a mail survey, the second wave sent Priority Mail, followed by telephone interviewing of nonrespondents	90 questions; 15 pages; complete redesign of content and format	Week of April 15, 1993	86% Missing item responses were also replaced via imputation
1995	Two waves of a mail survey, the second wave sent Priority Mail, followed by telephone interviewing of nonrespondents	95 questions; 17 pages; addition of work history module	Week of April 15, 1995	85% Missing item responses were also replaced via imputation

* Response rates between 1973 and 1989 are unweighted and defined as the number of eligible responses divided by the estimated number of eligible sample cases. Ineligible cases are those who are deceased, age 76 and up, residents of a foreign country, or institutionalized.

** Response rates between 1991 and 1995 are weighted and defined as the number of returned cases multiplied by their sample weights divided by the number of cases in the total sample multiplied by their sample weights.

- In 1991, the sampling frame was redefined in a major way, corresponding to the beginning of the SESTAT data system and the integration of the SDR with the National Survey of College Graduates and the National Survey of Recent College Graduates. In that year, individuals were included on the frame who were under 76 years of age, and foreign-earned doctorates and field-switchers were dropped from the frame. This change served to clarify the target population, which was henceforth defined as individuals under 76 years of age with a doctoral degree in a science and engineering field from a U.S. institution, who if foreign citizens, indicated plans to remain in the United States after receiving their degree.

2. Stratification Variables

- After 1973, the size of doctoral institution was dropped as a stratification variable because it did not correlate well with outcome variables.

- In 1975, minority status was added as a stratification variable, although it was known only for post-1973 doctorates, the year the question was added to the SED.

- In 1979, citizenship status was added as a stratification variable, although it was known only for post-1958 doctorates. In addition, the minority strata were collapsed and the number of fields was collapsed.

- In 1991, as part of the major sample redesign, the sample was completely restratified into about 240 cells, down from more than 2,000 cells in 1989. The stratification variables were redefined and included field of degree, gender, and demographic group—the latter a combination variable using disability status, race, ethnicity, and citizenship status.

3. Population Size

- The population of scientists and engineers who met the eligibility requirements for inclusion on the sampling frame grew from 261,393 in 1973 to 594,275 in 1995.

- This growth stems not only from increased Ph.D. production in those years, but also from expansion of the frame to include audiology and speech pathology (1983) and to accommodate a longer time frame (1987).

4. Sample Size

- Sample sizes and sampling rates fluctuated during the 25 years between 1973 and 1995, largely due to funding availability. The overall sampling rate was about 21 percent between 1973-1977. In 1979, the sampling rate was reduced to 12 percent and the sample size decreased from 68,532 cases in 1977 to 41,763 cases in 1979.

- The next change in sample size occurred in 1991. The overall sampling rate was reduced from about 14 percent to 7 percent, with a corresponding reduction in sample size from 73,611 to 37,999. This reduction was partially restored in 1993, when additional funds became available and the sample size was increased to 49,228.

- In 1995, a maintenance cut procedure was implemented to keep the sample size constant at around 50,000, regardless of the size of the population.

5. Data Collection Method

- From 1973-1989, the SDR was conducted as a mail survey. The number of mail waves varied between two and four. In some years, shortened questionnaires were used in follow-up mailings to encourage response.

- In 1991, computer-assisted telephone interviewing (CATI) was introduced to follow-up mail nonrespondents.

- The data collection protocol was changed again in 1993, when the number of mail waves that preceded CATI was reduced from three to two, and the second wave was sent Priority Mail. This protocol was followed again in 1995.

6. Questionnaire Content and Format

- The content and format of the questionnaire also changed over the years, as policy issues changed and advances in the field of survey methodology provided new means for addressing data quality. The questionnaire changes are too numerous to list here, but the overall pattern shows increasing questionnaire length— from 18 questions in 1973 to 95 questions in 1995.

- The format also changed as research became available suggesting formatting methods for reducing item nonresponse and improving response quality. In 1991, "respondent friendly" design principles were introduced to the mail questionnaire.

7. Reference Date

- In the first two rounds of the SDR, the survey asked questions about two time periods—the survey year and the prior year.

- In 1977, the week of February 6-12 was the reference period. From 1977-1989, the reference date was either a week in February or the month of February.

- In 1991, the first major shift occurred and the reference date was moved to September. This move was needed because the sample redesign in that year delayed the start of the project.

- In 1993, the reference date was shifted back to the week of April 15. This change was made to put the three surveys in the SESTAT system on the same schedule. The reference date remained April 15 in 1995.

8. Response Rate

- During the period the SDR was strictly a mail survey, response rates ranged from a high of 75 percent in 1973 to a low of 55 percent in 1989.

- When telephone follow-up interviewing was introduced in 1991, the response rate increased to 87 percent and remained close to that level in 1993 and 1995.

This summary of changes is provided in order to understand the environment in which the SDR has operated in the past and to provide a context for discussing the strengths and weaknesses of the current design in the next chapter. For more detailed descriptions of the 1989 to 1991 survey designs see Mitchell et al. (1998).

3. THE CURRENT SDR

Any design change contemplated for the SDR must begin with an assessment of the strengths and weaknesses of the current design. The 1995 survey is summarized in this chapter as the base from which proposed changes and their consequences are discussed. We profile the 1995 survey because it is the latest year for which a complete set of survey outcomes and population estimates are available. The 1997 survey was similar to the 1995 survey in the key features that relate to sample redesign issues. We conclude this chapter with a discussion of the strengths and weaknesses of the current design.

A. Design Attributes

When it was first established in 1973, the objective of the SDR was to profile the employment outcomes and attributes of the entire U.S. population of doctorate recipients in S&E. The target population was envisioned as consisting of (1) S&E doctorates from U.S. institutions, (2) non-S&E U.S. doctorates working in an S&E occupation, and (3) S&E doctorates from foreign institutions.[4] In the 1980s, though, it became clear that the SDR did not have access to an adequate frame for effective inclusion of S&E doctorates earned at foreign colleges and universities (Mitchell et al. 1998). Hence, SDR's target population is now restricted to individuals who received a doctorate in a S&E major field of study from an educational institution in the United States (or a U.S. territory) between January 1, 1942 and June 30th of the year preceding the survey. In addition, as of that survey's reference week, the doctorate recipient must have been living in the United States or its territories, noninstitutionalized, and age 75 or younger (Brown 1997).

SDR constructs its frame from the Doctorate Records File (DRF), a historical compilation of data from the Survey of Earned Doctorates (SED). The coverage of the SDR frame is dependent on DRF coverage, which in turn is dependent on SED coverage. Fortunately, the SED obtains at least commencement program data (name, degree field, date of degree conferral, etc.) from all graduating doctorate recipients. Multiplicity for individuals who receive more than one doctoral degree (an uncommon but not rare occurrence) is minimized because the DRF only includes a record for the first doctorate.[5] *Individuals with a second doctorate in an S&E field will be missing from the frame when their first doctorate is in a non-S&E field. Although the impact of this omission in the traditional SDR frame is unlikely to be substantial, dual doctorates should be examined and corrective action taken as appropriate.*

The first step in SDR sampling is establishing the requirements for inclusion in the sampling frame. In developing its frame, the SDR excludes those non-U.S. citizens who indicate in the SED that they have definite plans to leave the United States after receiving their doctoral degree. *This exclusion assumes that these individuals are correctly predicting their future behavior and that any undercoverage of the population of U.S.-earned S&E doctorates will be negligible. This assumption should be verified in light of recent SED methodological research to determine whether the correlation between definite plans and later actions is as high as might be desired for frame building.*

---

4 From 1977 to 1995, 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.

5 Subsequent questionnaires from multi-doctorate individuals are stored in hard copy form only.

While perhaps erring on the side of being too liberal in excluding non-U.S. citizens with intentions to leave the country, the SDR takes a rather conservative approach once an individual is selected for the sample. Once selected, individuals are considered "temporarily ineligible" when they leave the country. Every survey cycle thereafter, SDR locates the person and confirms that he or she is still living outside the United States.[6] While U.S. citizens often return from overseas residence, only a small percentage of non-U.S. citizens return after moving abroad. *From an undercoverage reduction standpoint, it might be more appropriate to allow non-U.S. citizens to be included in the frame for initial sample selection, regardless of their intentions, and then to label them as "permanently ineligible" if they move abroad. U.S. citizens should be followed and their residency established every cycle because they are much more likely to return. The SDR database can be used to validate this suggestion, which should be separately investigated for U.S. versus non-U.S. citizens.*

Historically, SDR has used a stratified sample design with strata reflecting shifting policy interests. The original sample was selected using stratified random sampling with strata based upon field of degree, year of degree, size of doctoral institution, and gender. The original plan was to continue to collect data biennially from these cases, and from the new cohorts added in each cycle thereafter. However, the design has been repeatedly modified (see Table 1). By 1995, the SDR design had become a complex, multi-phase design, with unequal weights within the last phase strata.

The most important SDR design change was the restratification of the sample and the frame in 1991. Since then, SDR strata have been defined based upon demographic group, major field of study, and gender. Demographic group is a composite variable that categorizes doctoral recipients in terms of their disability status, Hispanic origin, race, citizenship status, and birthplace (U.S. vs. foreign). The variables used to define demographic group were introduced into the SED at different times, with race only becoming available in 1975 and disability status becoming available in 1985. For doctorates graduating before 1958, for instance, only gender and field of study are available in the DRF. Survey responses were used for stratification in 1991 when DRF data were missing. Post-1991 doctorates were classified directly into the new strata using frame information only. For pre-1991 doctorates, the current SDR is treated as a two-phase sample design, where the Phase 1 strata are the strata used prior to 1991 and the Phase 2 strata are the current strata defined using pre-1991 survey responses.[7] In the absence of information, the 1991 SDR and later cycles assumed that the sampling weights were equal within the 1989 strata and that the 1989 design could be regarded as a stratified, simple random sample. Within these Phase 2 strata, the sampling weights vary depending on the person's pre-versus-post 1991 selection status and (if pre-1991) the Phase 1 strata for which he or she was originally selected.

The traditional SDR sampling approach is to select a sample of new doctorate recipients for each cycle, which is then added to the previous cycle's sample; sample members who are deceased or 76 years old and older are dropped. The intent of SDR sample selection procedures has been to select samples using the same selection probabilities within strata regardless of the cycle in which the selection was added to the SDR. Once selected for the SDR, sample cases have tended to be retained indefinitely, except for sample size reductions associated with budget cuts.

---

[6] In the 1997 SDR, persons who were residing abroad during the previous survey cycle were automatically assumed to be living abroad in 1997, regardless of citizenship.

[7] The pre-1991 SDR sample members are actually derived from a multi-phase design with unknown and (possibly) unequal sampling weights within the last phase strata. See Mitchell et al. (1998) for further details.

In recent years, however, the overall sample size has been fixed at about 50,000, with a "maintenance" cut taken of the previous cycle's old sample, which is then added to a freshly selected sample of recent doctorate recipients. Because of this maintenance cut and the introduction of new stratification variables in 1991, the current SDR design is quite complex with each unit's overall probability of selection the product of its initial selection probability times the various probabilities with which the sample unit has been subsampled after its initial selection.

The 1995 SDR sample design divided the target population into three groups:

1. The pre-1991 population (degrees received prior to July 1, 1988) represented by SDR selections made prior to the 1991 SDR for whom 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).

2. 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.

3. 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 2 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 sampling weight (or a size measure of 1 for substratum 3 selections sampled directly from the DRF). This PPS selection made the 1995 sampling probabilities less variable within strata and hence reduced the deleterious effect of unequal weighting associated with 1993 sample selections (but added an unequal weighting effect associated with rounding to create integer substratum sample sizes).

SDR precision differs from that of simple random sampling due to the extensive oversampling across Phase 2 strata (and attendant across-strata unequal weighting effect) and the unequal weighting within Phase 2 strata resulting from the reclassification of the more than 2,000 Phase 1 strata from the pre-1991 design (with varying selection probabilities and weights) into the current Phase 2 strata. Many Phase 2 strata have very small sample sizes and very small population sizes. Most problems are associated with rare populations such as the disabled or Native American doctorates, who are further partitioned by major field of study and by gender for stratification purposes. Collapsing the strata for such rare subgroups would improve the precision of analyses for these rare populations, without adversely affecting precision for analyses not involving these variables.

We should also note that the demographic group stratification variable does not conform to the way researchers group data in analysis. U.S. citizens are stratified by race and ethnicity, while foreign-born doctorates are stratified by citizenship status. Analysts tend to analyze racial and ethnic groups without partitioning them based upon citizenship. For estimates of black doctorates, they combine sample cases from these four different demographic groups, each with different selection probabilities: (1) disabled blacks, (2) U.S.-born blacks, (3) naturalized blacks, and (4) foreign blacks. It would be preferable to group blacks together, with the possible exception of the U.S.-born disabled. (The U.S.-

born disabled are sampled at a higher rate than U.S.-born blacks who in turn are sampled at a higher rate than blacks who were naturalized or foreign citizens at the time of degree award.) Citizenship is another important analysis domain that can differ from the stratification variable. Citizenship at the time of degree award may not show sufficient correspondence to current citizenship status to merit the current strategy of distinguishing between naturalized, foreign-born doctorates and other foreign-born doctorates in creating strata.

The *exact* probabilities of selection are unknown at present and determining the exact values might be difficult. Through the 1980s, the probability of selection was not recorded on the SDR database. Instead, SDR calculated the analysis weight for respondents within stratum h as the estimated stratum population size of eligible cases $\hat{N}_{pe}(h)$ divided by the number of responding, eligible sample cases $n_{re}(h)$. Stratum h eligible cases were estimated as the product of the stratum population size $N(h)$ and the proportion of eligible respondents to total respondents. Identified permanent ineligibles (e.g., those dead or over 75) and identified temporary ineligibles (e.g., those overseas or institutionalized) were considered to be responding but ineligible. With equal probabilities of selection within strata, this weighting approach is reasonable and can be shown to result from the assumption that nonresponse occurs at random within strata. However, as Table 1 demonstrates, SDR strata have not remained constant. When stratum changes occurred, SDR reclassified the older cohorts into the new strata. However, the weighting plan remained unchanged with the analysis weight defined based upon the new strata. Recovering the exact probability of selection for cases initially selected prior to the 1990s will be difficult as the data are stored on out-of-date media. In constructing the weights for 1991 on, SDR assumed that the sampling weights for pre-1991 initial cohorts could be treated as equal within strata. This assumption should be at least approximately correct.

SDR sample allocation appears to be far less than optimal. Both within and across strata, there is extensive unequal weighting that does not reflect appropriate oversampling of small populations. Many strata are too small to serve effectively as strata. For example, one 1995 stratum has a population size of four and a sample size of one. Variable definitions used for stratification also differ from the definitions routinely used in analysis. For the data analyst, this design inefficiency translates into wasted observations; sample sizes for a subpopulation that appear more than adequate may result in insufficient precision for sample estimates.

Inclusion in the SDR sample is virtually a "life sentence." Once selected a sample person tends to stay in until death or age 76.[8] The original 1973 sample selections have been interviewed as many as 12 times. With an average life span now of more than 75 years, a member of the SDR sample could be interviewed 25 times before becoming ineligible. We are aware of no other longitudinal survey that interviews sample cases this often.

Repeated interviewing of the same individuals may be leading to declining response rates, with sample loss occurring disproportionately during the middle years of the doctorate's career. *An investigation into this possibility of differential sample loss is needed. That investigation should also examine the extent to which hard-core refusals may be growing.* In 1995, eight percent of the sample were designated as "permanent refusals" (sample cases who are *never* to be contacted again), up from five percent in 1991.

[8] Funding constraints and sample design changes have caused some cases to be dropped without being added back in later years. In 1995, substratification into three cohort groups was added to facilitate a maintenance cut to force the sample size to the 1993 total of about 50,000 cases.

Any such investigation of a time-in-sample effect should also determine if the sample distribution is being adversely affected by design changes made in the 1990s. Prior to 1991, SDR had more than 2,000 strata. The very large number of strata resulted from the decision to establish separate strata each time new cohorts were added. Other variables used for stratification in 1989 included field of degree, gender, minority status, and citizenship. Following the advice of the Committee on National Statistics (Citro and Kalton 1989), SDR eliminated cohort as a stratification variable beginning in 1991.[9] Because SDR has tended to use the strata to define weighting classes, these changes in strata could also result in differences in the survey estimates when differential patterns of nonresponse are associated with changes in stratum definitions. From 1991 on, though, four age categories were used in addition to strata in defining weighting classes, which may ameliorate some of the potential biasing effects associated with the changes in stratification/classing procedures between pre-1990 and post-1990 estimates. If response rates vary by cohort, the current sample design may still exhibit some time-in-sample bias after 1991.[10]

The current SDR design is envisioned as a longitudinal survey of U.S.-earned S&E doctorates, with a sample of new cohorts added every two-year survey cycle and the previous samples of older cohorts followed for interview when they meet eligibility criteria. Under this vision, the SDR would provide a life history of the employment attributes and outcomes for the nation's U.S.-trained S&Es. Changes over time could be studied and related back to degree, demographic characteristics, and activities after degree award. Unfortunately, the current SDR has design and operational flaws that make it difficult, if not impossible, to analyze the resultant data as a longitudinal design. Longitudinal designs freeze the sample at the time of initial selection and follow it without change thereafter. The SDR sample design for new and continuing cohorts has been modified multiple times over its 25 year history, with changes in stratification, selection probabilities, sample sizes, and so forth. In addition, the 1991 SDR had a major design contraction; the sample size went from 73,611 in 1989 to 37,996 in 1991 and then to 49,228 in 1993.

By definition, longitudinal analyses require repeated measurements over time from the same individuals. Nonresponse is always a problem for survey analysis, but particularly so for longitudinal analyses, which require responses from more than one survey cycle. For this reason, longitudinal surveys typically set goals for and monitor the extent of sample retention over time, as well as establishing goals for response rates for the individual survey cycles. In addition, longitudinal designs incorporate a data collection approach that yields high response rates to the individual survey cycles, to ensure reasonable response rates for the entire longitudinal data series. Up until 1991, the SDR was a mail survey only. While response rates were high compared to other mail surveys, they were not as high as desirable for longitudinal database construction. In 1989, the last year with a mail-only design, the SDR achieved an overall response rate of 55 percent. To improve response, SDR began using telephone followup to collect data from mail nonrespondents. Overall response rates increased to 87 percent in 1991 and 86 percent in 1993. Unfortunately, the longitudinal nature of the data series was affected in 1991, 1995, and 1997 by funding constraints that were insufficient to allow telephone followup of all mail nonrespondents.

---

[9] Other changes were to use a more detailed racial/ethnic classification and to add disability status as a classifying variable. In 1995, substratification into three cohort groups was added to facilitate a maintenance cut to force the sample size to the 1993 total of 50,000 cases.

[10] This observation is reinforced by the change in estimates associated with race/ethnicity observed between 1989 and 1991 when the previous minority/nonminority stratification variable was broken down by racial/ethnic group in sampling and weighting. (See Mitchell, et al. 1998 for details.)

Finally, longitudinal analyses require the development of specialized longitudinal weights for use in analysis. The number of separate, longitudinal weight variables that are needed is often as many as there are combinations of the various survey cycles. Thus, for a longitudinal survey with three survey cycles, a common situation is to develop analysis weights for respondents to Cycle 1, Cycle 2, and Cycle 3 individually; then analysis weights for respondents to both Cycle 1 and 2, to both Cycle 1 and 3, and (possibly) to both Cycle 2 and 3; and finally for respondents to all three surveys.

To illustrate why many analysis weights may be needed, suppose an analyst is interested in (1) the percentage of new doctorates that accept initial positions in academe and continue employment in academe and (2) in how this statistic has changed over time. Operationally, the analyst might take each survey cycle's new cohort sample and restrict their analysis to those that had accepted academic positions. Then the analyst might merge in data from the same individuals 10 years later (i.e., five survey cycles later).[11] For their data record to be used in this analysis, a new doctorate must respond to the initial interview and to the one 10 years later. To analyze the combined data set, the analyst needs a nonresponse-adjusted analysis weight that requires that the individual "respond" to both surveys in order for him/her to be labeled as a respondent.[12] Creating this longitudinal weight requires that the probability of selection into both surveys (the initial one when the doctorate is a new cohort and the one 10 years later) be known. The frequent changes in the SDR design and the use of subsampling in nonresponse followup might make it difficult or intractible to determine the probability with which sample units were included across multiple survey cycles. Calculating a longitudinal weight and adjusting it for nonresponse requires knowing which sample cases should have responded in the time periods under study and what the probability of joint inclusion is for these cases.

To the best of our knowledge, SDR longitudinal weights have not been developed, and appropriate longitudinal analyses have not been conducted with the SDR. The SDR is frequently used for analyses across time, however. These time series analyses tend to treat the SDR as a repeated, cross-sectional design and to compare estimated totals, means, and proportions across time, where each estimate is calculated using the analysis weight for that survey cycle. These analyses only require use of an appropriate analysis weight for each survey cycle that recognizes the probability of selection for that survey cycle and that adjusts for nonresponse to that survey cycle. Such analyses do not require a longitudinal design, but comparisons across time benefit from the correlation associated with overlapping samples over time.[13]

B. Strengths and Weaknesses

We view the primary advantage of the present data series to be the fact that it has followed individuals from degree receipt (until age 76), allowing the analyst to study career progression and retirement rates and patterns. Unfortunately, this capability has not been fully developed. To implement statistically valid analyses, longitudinal weights need to be developed that reflect the probabilities of inclusion across rounds and that adjust for nonresponse. Such weights will be difficult to develop due to

---

[11] Because some doctorates may die, become institutionalized, or leave the country during this time period, the analysis might be restricted to individuals who are noninstitutionalized and living in the U.S. 10 years later. The required information is available for survey respondents.

[12] Individuals identified as ineligible for interview would be considered to have responded in making this adjustment.

[13] SDR reports provide generalized variance functions that allow the analyst to estimate the standard errors for survey estimates. However, they do not provide a mechanism for determining the standard error of changes from one time period to another; the standard error of such change estimates includes a covariance term for which SDR reports do not provide information.

the sample variations that have occurred over time. The number of cases available for longitudinal analysis may be small due to the lower response prior to 1991, the reductions in sample sizes periodically made, and the use of maintenance cuts to control overall sample size.

Another advantage of the current design is that the sample size is sufficient to allow time series analyses by broad fields of study. Further, the overlaps between the samples over time periods results in more accurate estimation of change over time than would be produced by independently selected samples for each time period. The use of PPS subsampling for the maintenance cuts is also reducing the unequal weighting within strata over time.

Still, the present design is far from optimal. A redesigned sample could increase precision for survey estimates by improving the sample allocation and creating strata that correspond more closely to reporting domains. It is also troubling that the probabilities of selection can only be approximated for the current sample. Finally, the level of burden that SDR is imposing on survey respondents is not justified, as the longitudinal aspect of the data set is not being used properly due to the absence of longitudinal weights. We suspect that nonresponse will increase over time as respondents notice SDR's "life sentence."

4. THE REFRESHED DESIGN

Continuing data series need to be periodically examined to determine if they are meeting current policy needs effectively. This examination is particularly important for a data series like the SDR which has now been in operation for 25 years. The design problems of the current SDR are the result of repeated "patches" to bring it into alignment with policy needs and funding availability. One way to remove these design inefficiencies and anomalies is to "refresh" the current design by selecting a new longitudinal sample to replace the current one. In this chapter, we describe issues in the creation and implementation of a refreshed design.

A. Statistical Design

Redesigns of continuing data series must be implemented in such a way as to minimize the impact on the time series. One way to do that is to gradually phase in the new sample. For instance, the SDR might be "refreshed" by replacing 20 percent of the sample in each of the next five survey cycles. After 10 years, the current SDR sample will have been totally replaced. This refreshment would

- improve the correspondence between sampling strata and reporting domains and thereby improve the precision for domains related to stratification variables,
- eliminate excessive inequities in sampling rates and thereby increase the overall precision of survey estimates for larger, cross cutting domains, and
- restratify the sample using frame variables available for each cohort (as opposed to ad hoc mixtures of frame and sample-derived responses).

Again, by "refreshed design," we mean a new, freshly designed and selected, longitudinal sample phased in over a specified time period, say five survey cycles. Once fully in place, this sample would be followed for an indefinite time period, in much the same way that the current design is followed, with new cohorts added each survey cycle and cohorts over 75 deleted.

Phasing in a new longitudinal sample is somewhat complicated by the fact that SDR brings in a new cohort sample each year. To begin the process, we suggest that a new sample be selected equivalent to 100 percent of the current sample and that this "refreshed" sample be randomly allocated to five "fresh" replicates. This allocation would be done so that the selection probabilities are exactly equal within strata across the fresh replicates. Next, the current sample would be partitioned into five "old" replicates. Because of the wide variation in sampling weights within strata for the current sample, we recommend that the current sample be partitioned into five old replicates using PPS sampling, with the current sampling weight used as the size measure. PPS sampling will remove some of the inefficiency of the current design and improve overall precision.

In the first survey cycle, a new-cohort sample (for doctorates received in the last two academic years prior to the Cycle 1 reference year) would be added to the five fresh replicates and the five old replicates by selecting twice the appropriate full sample size of Cycle 1 new cohorts, partitioning that sample into 10 replicates, and then randomly assigning them to the five fresh replicates and five old replicates so that the selection probabilities are exactly equal for each replicate's new-cohort sample. At this point, each replicate will contain a valid national sample of doctorate S&Es through Cycle 1's cut-off date for degree receipt. Fresh replicate 1 would be included in Cycle 1 and all cycles thereafter. Old replicate 1 would be dropped from the sample in Cycle 1. Thus, the Cycle 1 sample would be composed of fresh replicate 1 and old replicates 2, 3, 4, and 5.

In Cycle 2 this process would continue. First, twice the desired sample size of Cycle 2 new cohorts would be sampled and then allocated randomly to the five fresh and five old replicates. Then, the Cycle 2 sample would be constructed as fresh replicates 1 and 2 and old replicates 3, 4, and 5.

In Cycle 3, twice the desired sample size of Cycle 3 new cohorts would be sampled and then allocated randomly to the five fresh and five old replicates. Then, the Cycle 3 sample would be constructed as fresh replicates 1, 2, and 3 and old replicates 4 and 5.

In Cycle 4, twice the desired sample size of Cycle 4 new cohorts would be sampled and then allocated randomly to the five fresh and five old replicates. Then, the Cycle 4 sample would be constructed as fresh replicates 1, 2, 3, and 4 and old replicate 5.

By Cycle 5, no old replicates would be included in the SDR. The sample would be composed of the five fresh replicates, plus the required sample of new Cycle 5 cohorts.

In implementing the refreshed design, the stratum definitions would need to be reconsidered. A symmetric approach has been used that defines strata based upon all possible crosses of demographic group, degree field, and gender and then applies separate precision constraints for each stratum. This approach produces strata with adequate population sizes for whites, but for rarer domains (e.g., the disabled, Native Americans) results in very small strata. These strata can be so small that separate estimation for them would violate respondent confidentiality. A minimum population size needs to be established for strata and appropriate collapsing preformed based on policy implications. For example, Native American, female, computer programmers is one small stratum in current use. Here, the decision might be to drop gender as a stratification variable and, if needed, to collapse degree field into larger categories (e.g., combine computer programmers with mathematicians and statisticians into a group called "mathematical and information sciences"). If the resultant stratum is still too small, one might drop the requirement to stratify by degree field altogether.

In selecting the replicate samples from these new strata, care must be taken to avoid sampling the same individual for inclusion in more than one refreshed replicate. Such overlap between the refreshed replicates can be avoided by using a permanent random number technique in combination with stratified simple random sampling (see Ohlsson 1995). With this approach, a permanent random number would be generated for each Doctorate Records File (DRF) record. Within strata, the records would be sorted by this permanent random number. Then the replicate 1 sample of size $n_1(h)$ from stratum h would be selected by taking the first $n_1(h)$ records from the randomly ordered list for stratum h. The replicate 2 sample of size $n_2(h)$ would be selected by taking the next $n_2(h)$ records from the ordered list, and so on.

In implementing this refreshed design, care also must be taken to avoid minor weighting variations within strata, say across replicates or between new cohorts and old cohorts. This problem is manifest in the current design and results from the need to translate a desired sample size allocation that is a real number (e.g., 32.45) into the integer value needed for selection purposes (e.g., 32). The usual rounding approach (round down if fraction less than 0.5; round up otherwise) results in substantial amounts of unequal weighting for the current design and might prove much worse for the refreshed design. We recommend the use of a randomized rounding approach that enforces a fixed stratum sample size while producing exactly equal sampling rates within strata.

Another issue that must be addressed is SDR's use of stratification variables that are unknown for many—sometimes most—of the records in the DRF, which serves as SDR's frame. Availability is dependent upon when the Survey of Earned Doctorates (SED) began collecting the information. Table 1 shows when SDR began including variables such as citizenship (1979), disability status (1991), and detailed racial category (1991). SED collected these variables before SDR introduced them as stratification variables, with citizenship collected in 1958, for instance. There is no effective, yet relatively inexpensive way, to oversample for characteristics for which the required frame variables are unavailable for most doctorates.

An example is the disabled. To oversample the disabled, the most effective method in terms of sample precision would be to select a larger-than-needed sample of older cohorts for whom disability data are unavailable and then screen the selections to determine their status. Those who said in screening that they were disabled would be taken in their entirety while the nondisabled would be subsampled. Obviously, this method requires more interviewing effort and is more difficult to implement in a mail survey. The alternative to screening is to redefine study objectives. Thus, NSF might restrict its specialized analyses to doctorates received during the years when SED collected the required stratification data. Under this scenario, analyses for the disabled would have to be restricted to disabled doctorates receiving degrees after 1990. For other groups, the situation is not as bad. SED has collected detailed race classifications since 1973. The stratification plan also needs to reflect the available data. We envision strata that reflect the changes in data availability. The availability for the first set of strata for the oldest doctorates might be based upon degree field, gender, and age. (Age is added here for stratification to control for the age ineligibility of the sample as time passes.) Then the next set of doctorates in time might be stratified by race/ethnicity, degree field, and gender. The third set of doctorates in time might be defined based upon citizenship, race/ethnicity, degree field, and gender. The most recent doctorates might be stratified by disability status, race/ethnicity, citizenship, degree field, and gender.

In the last decade, funding constraints have caused the SDR to contract in 1991, then expand in 1993, to introduce maintenance cuts in 1995 (to control the overall sample size to a fixed number when new cohorts are added), and to unexpectedly subsample mail nonrespondents in 1995 and 1997. Unless nested in a long-range plan, such variations in the sample can harm the longitudinal nature of the design. *We suspect that such funding uncertainties are an economic fact of life, but that NSF may also desire a longitudinal design capability. In which case, we suggest that NSF consider committing itself to a smaller longitudinal sample that is guaranteed for automatic interview by mail and telephone at each cycle and for which longitudinal weights will be developed.* That sample could then be supplemented with extra selections that can expand and contract in response to funding availability or that can be subjected to nonresponse subsampling if needed. Separate weights would be developed that combined the longitudinal sample with the extra selections for cross-sectional analysis. The supplementation of the longitudinal sample would improve the precision of cross-sectional and time series estimates, which now form the bulk, if not all, of SDR analyses.

B. Impact on Survey Operations

There are two impacts on survey operations to consider if the refreshed sample design is implemented. The first relates to locating new sample members. Because sample members will be selected who are up to 75 years old, the permanent address information they supplied on the SED questionnaire can be quite out of date. Further, some of the oldest cases may not have completed an SED questionnaire, which was not used until 1958. The presence of social security numbers (SSNs) for the large majority of individuals on the frame, however, will ameliorate the locating problem.[14] The availability of SSNs makes electronic database searches a fast and inexpensive method for obtaining current address information. For those without SSNs, a variety of custom locating techniques are available, for example, name and address searches using the Internet. The SDR's history shows that locating rates over 90 percent can be expected with this population, given a sufficient commitment of time and resources.

The second potential problem relates to the availability of "base" data, or background information from the SED, for newly selected sample members. Because skeletal records were created for some DRF cases, and because the SED did not ask some demographic questions until later in its timeline, there will be missing base data for some sample cases. This missing data will include variables such as the institution and year for the bachelor's degree, the highest level of parents' education, citizenship at the time of degree award, and race. These questions of important analytic interest are not now asked on the SDR questionnaire because they have been collected over the years for the large majority of cases. This means that a separate mail questionnaire will need to be developed to collect this information from the newly selected cases. While not an onerous task, working with two questionnaires will add complexity to the survey mailings. Of course, with CATI administration, these questions can be programmed in advance and asked only if the data are missing.

C. Strengths and Weaknesses

The primary advantage of the refreshed design is that it contains almost all the advantages of the current design while eliminating almost all its disadvantages. The current design is far from optimal in its sample allocation, resulting in loss of precision due to excessive weighting variations, both within and across strata. With a refreshed design, the sample could be redesigned using up-to-date sampling procedures. Unlike the current design, the exact probabilities of selection would be known for the refreshed portion of the design. The refreshed design would be longitudinal in nature so that a new longitudinal data series would be gradually brought in to replace the current design. Because the current design would be phased out gradually, the effect on the time series would be minimized. To the extent that SDR's repeated interviews are having a deleterious effect on response, the refreshed design may increase response rates as it reduces overall burden on individual respondents.

The disadvantage of the refreshed design is mainly that it would end the current longitudinal data series. As we noted earlier, though, we see little evidence that SDR's longitudinal potential is being realized and are unaware of users who have constructed the specialized longitudinal weights needed for statistically valid analyses. When SDR was redesigned in 1991, sample information was used for stratification purposes—primarily to define disability status, race, citizenship, and so forth when these data were not collected in the SED. This stratification ability would be unavailable for the refreshed design. How much of a loss this is depends upon how effective the current sample is in capturing groups like the disabled. Although the actual sample count for domains like the disabled and Native Americans

[14] While the SED has always asked for SSNs, some item nonresponse occurs. In addition, Social Security numbers tend to be unavailable prior to 1958.

may be suitably large, once the design effect associated with unequal weighting is factored in we suspect the effective sample size will be substantially reduced and that many analyses will not be feasible.

5. THE ROTATING DESIGN

The rotating design is commonly used for continuing data series such as the SDR. Rotating designs split the sample into replicates. At each survey cycle, a new replicate is brought in and the oldest replicate is retired. Rotating designs achieve the advantages that sample overlap has in reducing costs and in increasing precision while limiting burden on respondents and any deleterious effects associated with indefinite followup. This chapter discusses the statistical and operational issues associated with converting the SDR into a longitudinal sample.

A. Statistical Design

A rotating design for the SDR would begin the same as the refreshed design discussed in Chapter IV. Assuming the decision was made to keep each sample replicate in for ten years, a new replicate of about 20 percent of the total sample allocation would again be brought in every survey cycle, while the equivalent amount of old sample would be retired. The designs would remain identical for the first five survey cycles. The difference between the refreshed and rotating designs occurs after the fifth survey cycle. From the sixth survey cycle on, the refreshed design follows all of the refreshed selections, while adding a new cohort sample to be followed indefinitely. The rotating design drops the oldest replicate in the sixth survey cycle, selects a new replicate sample from the old cohort population, and then adds a new cohort sample to each of the resultant five replicates. This process is repeated for each survey cycle thereafter, with the oldest replicate being removed from the sample, a new replicate added, and a new cohort sample added to the resultant five replicates.

For the rotating design, it is again important to choose an approach that prevents unequal weighting effects across replicates and that minimizes the probability that individuals will be repeatedly sampled. The procedures described for the refreshed design are pertinent here as well. In addition, each stratum's records should be considered to be a circular list so that when the list is exhausted, the process continues with the first record in the randomly ordered list for stratum h.

B. Impact on Survey Operations

The impact on survey operations is the same for the refreshed and the rotating design for the first ten years. Thereafter, the refreshed design is followed much as the static current SDR is followed. For the rotating design, however, each cycle brings a new sample panel that needs tracing and locating and that needs base information collected.

C. Strengths and Weaknesses

The rotating design results in an efficient sample design for cross-sectional analyses with known probabilities of selection. It retains sufficient sample units across cycles to allow for efficient comparisons over time, as long as the estimates being compared are for survey cycles in reasonably close proximity. For survey cycles separated by 10 or more years, there would be no shared sampling units. By rotating the sample out after 10 years, the burden placed on individuals is reduced. Reducing the burden may make data collection easier by reducing refusals and it may improve data quality.

The weakness of the rotating design is that it will support longitudinal analyses only for 2 to 10 years with ever decreasing sample sizes as the years between the beginning and end data increase. For two years apart, 80 percent of the sample would be available for longitudinal analyses; thereafter the longitudinal sample would decline to 60 percent for four years apart, 40 percent for six years apart, and 20 percent for eight years apart. How significant a weakness this is depends upon the importance of being able to do longitudinal analyses over long periods of time with large sample sizes. The analyses over time that appear in SDR reports tends to be time series analyses rather than longitudinal analyses. While time series analyses benefit precision wise from shared samples over time, overlapping samples are not required.

6. CONCLUDING REMARKS

Which of these three designs is the best design for the future is a policy decision for NSF to make. Much of the decision revolves around the question, "How important are longitudinal analyses to NSF?" One way to answer this question is to look at what analyses over time have been conducted in the past and how analyses are expected to differ in the future. We have not identified any correctly executed longitudinal analyses, but we have found numerous instances of time series analyses which benefit from shared samples but do not require them.

As our last comment, we emphasize that the refreshed design and the rotating design are exactly the same for the first 10 years. Even after these 10 years, if longitudinal analyses should become important, the rotating design can be converted back into a longitudinal design.

The most important decision for NSF is whether the current longitudinal ability is so important that it is vital to retain the current inefficient, statistically compromised design. If the answer is "yes," then we recommend further research to address the design flaws of the current design, including an investigation into whether the probabilities of selection can be resurrected from the historic software language, media, and computers used for storage and whether longitudinal weights can be developed.

REFERENCES

Brown, P. "Methodological Report of the 1995 Survey of Doctorate Recipients." National Research Council, 1997.

Citro, C.F. and G. Kalton (eds.). *Surveying the Nation's Scientists and Engineers, A Data System for the 1990s*. Washington, DC: National Academy Press, 1989.

Ohlsson, E. "Coordination of Samples Using Permanent Random Numbers." In B. G. Cox, D. A. Binder, B. N. Chinnappa, A. Christianson, M. J. Colledge, and P. S. Kott (eds.), *Business Survey Methods*, New York: John Wiley & Sons, Inc., 1995

Lindsey R. Harmon, "Sampling Scheme for Comprehensive Doctorate File." Memorandum to W.C. Kelly, National Research Council, 1972.

Mitchell, S. B., R. Moonesinghe, and B. G. Cox. "Using the Survey of Doctorate Recipients in Time Series Analysis: 1989-1993." Final report submitted to the National Science Foundation under a subcontract to the National Research Council. Washington, DC: Mathematica Policy Research, Inc.

National Research Council. "Questionnaire Survey of Comprehensive Roster of Doctoral Scientists and Engineers." Washington, DC: National Academy of Sciences, 1972.

National Science Board. *Science & Engineering Indicators*—1996. Report No. NSB 96-21, Washington, DC: U.S. Government Printing Office, 1996.

National Science Foundation Web Site. "About the NSF." [http://www.nsf.gov/home/about/creation.htm]. May 6, 1998.