Improving the Design of the Scientists and Engineers Statistical Data System (SESTAT)

Committee to Review the 2000 Decade Design of the Scientists and Engineers Statistical Data System (SESTAT), National Research Council

ISBN: 0-309-50413-9, 48 pages, 6 x 9, (2003)

Improving the Design of the

Scientists and Engineers Statistical Data System (SESTAT)

Committee to Review the 2000 Decade Design of the
Scientists and Engineers Statistical Data System (SESTAT)

Committee on National Statistics

Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract/Grant No. SES-0112521 between the National Academy of Sciences and the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-087112

Additional copies of this report are available from National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

Suggested citation: National Research Council. (2003). *Improving the Design of the Scientists and Engineers Statistical Data System (SESTAT).* Committee to Review the 2000 Decade Design of the Scientists and Engineers Statistical Data System (SESTAT). Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE TO REVIEW THE 2000 DECADE DESIGN OF THE SCIENTISTS AND ENGINEERS STATISTICAL DATA SYSTEM (SESTAT)

ROBERT BELL (*Chair*), AT&T Laboratories-Research, Florham Park, NJ
BARBARA BAILAR, National Opinion Research Center, University of Chicago (emeritus)
PAUL BIEMER, Research Triangle Institute, Research Triangle Park, NC
BENJAMIN KING, Statistical Consultant, Del Ray Beach, FL
ROBERT SANTOS, NuStats, Austin, TX
LOWELL J. TAYLOR, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University

JAMES P. MCGEE, *Study Director*
MARIA ALEJANDRO, *Project Assistant*

COMMITTEE ON NATIONAL STATISTICS
2002

JOHN E. ROLPH (*Chair*), Marshall School of Business, University of Southern California
JOSEPH G. ALTONJI, Thomas DeWitt Cuyler Professor of Economics, Yale University
ROBERT BELL, AT&T Laboratories-Research, Florham Park, NJ
LAWRENCE D. BROWN, Department of Statistics, University of Pennsylvania
ROBERT M. GROVES, Survey Research Center, University of Michigan, Ann Arbor
HERMAN HABERMANN, United Nations Statistical Division, New York, NY
JOEL L. HOROWITZ, Department of Economics, Northwestern University
WILLIAM KALSBEEK, Survey Research Unit, Department of Biostatistics, University of North Carolina
ARLEEN LEIBOWITZ, School of Public Policy and Social Research, University of California Los Angeles
THOMAS A. LOUIS, Department of Biostatistics, Johns Hopkins University
VIJAYAN NAIR, Department of Statistics, University of Michigan, Ann Arbor
DARYL PREGIBON, AT&T Laboratories-Research, FlorhamPark, NJ
NORA CATE SCHAEFFER, Department of Sociology, University of Wisconsin-Madison
MATTHEW D. SHAPIRO, Department of Economics, University of Michigan, Ann Arbor

ANDREW A. WHITE, *Director*
DANELLE DESSAINT, *Senior Project Assistant*

Preface

Although representing a small fraction of the U.S. labor force (approximately 12 million individuals, about 8 percent of the labor force), scientists and engineers have major effects on the economic development of the country and on the rapid technological change that characterizes American society. Consequently, policy makers and researchers in government, industry, and academia need timely information about the numbers and characteristics of scientists and engineers in the United States. The key repository of that information, the Scientists and Engineers Statistical Data System (SESTAT), was created by the U.S. National Science Foundation (NSF) a decade ago and has been maintained by that agency. NSF has also maintained the primary sources for such information for decades.

The past decade has demonstrated the utility of SESTAT, but the SESTAT design shows some deficiencies with respect to response rates, coverage of populations of interest, and its ability to support some useful analyses. To tackle those deficiencies, NSF has proposed three possible design options for improving the database and asked the National Research Council's Committee on National Statistics (CNSTAT) to form the Committee to Review the 2000 Decade Design of the SESTAT.

This is the report of that committee. It presents our understanding of the purposes and characteristics of the SESTAT, applies the criteria we believe are important for assessing design options for the database, provides our recommendation for the best approach to adopt in the 2000 decade,

and offers our encouragement to NSF to pursue opportunities to improve the understanding of the numbers and characteristics of scientists and engineers in the United States.

We have been able to prepare this report because of the excellent cooperation and information obtained from the staff of the National Science Foundation and, particularly, its Science Resources Statistics (SRS) division. We are especially grateful to Norman Bradburn, director for the Social, Behavioral, and Economic Sciences Directorate; Lynda Carlson, director of SRS; and Ron Fecso, chief statistician of SRS, for providing to us information essential to our deliberations and for supporting our conduct of a workshop for wide discussion of many of the issues covered in this report.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Research Council (NRC). The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We thank the following individuals for their participation in the review of this report: Daniel Black, Center for Policy Research, Syracuse University; James M. Lepkowski, Institute for Social Research, University of Michigan; Fritz Scheuren, Statistics and Methodology, National Opinion Research Center; and Paula Stephan, Department of Economics, Georgia State University.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Lawrence Brown, Department of Statistics, University of Pennsylvania. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

This report is the collective product of the entire committee, and each member took an active role in drafting sections of chapters, leading discus-

sions, and reading and commenting on successive drafts. Staff at the National Research Council made important contributions to our work in many ways. We express our appreciation to Andrew White, CNSTAT director, for his valuable insight, guidance, and support; to Constance Citro, CNSTAT staff officer, for her invaluable intellectual support of the committee's work, including reviewing successive drafts of the report; and to Maria Alejandro, the panel's project assistant, who was indispensable in organizing meetings, arranging travel, compiling agenda materials, coordinating with the interested community, and managing the exchange of documentation among the committee members. We are deeply indebted to Eugenia Grohman, who significantly improved the report by dedicated application of her extraordinary editing skills.

Robert M. Bell, *Chair*
James P. McGee, *Study Director*
Committee to Review the 2000 Decade Design of the
Scientists and Engineers Statistical Data System (SESTAT)

Contents

Improving the Design of the

Scientists and Engineers Statistical Data System (SESTAT)

Executive Summary

SESTAT (the Scientists and Engineers Statistical Data System) is a system of surveys, developed and maintained by the Science Resources Statistics (SRS) division of the National Science Foundation (NSF) to provide timely information about the numbers and characteristics of scientists and engineers in the United States.

The committee examined three options presented by SRS to improve the SESTAT design: (1) conduct a new National Survey of College Graduates based on the 2000 census frame; (2) continue the existing 1990s panels; and (3) adopt a hybrid option, combining features of the 2000 census frame and the 1990 census frame options. Our primary criteria for evaluating these options were how well the design achieves and maintains an adequate response rate, how well it covers the complete population of interest, and whether sample sizes permit sampling precision that is adequate for the principal uses of the survey data. We also considered how well the designs support analyses of longitudinal data, trends, and biases. The committee was not asked, nor did it attempt, to perform an analysis of cost and implementation factors associated with various design options for SESTAT, examine the content of the surveys that support SESTAT, or revisit the question of how the population of scientists and engineers should be defined.

We conclude that the 2000 census frame option is the best choice of the three design options. We offer three primary recommendations:

Recommendation 1: Almost all of the resources allocated to the SESTAT data collection effort in 2003 should be devoted to drawing a new National Survey of College Graduates from the 2000 census and supplementing this panel with the National Survey of Recent College Graduates.

Recommendation 2: If SRS staff confirm that a targeted sample could be useful for the purpose of adjustment, SRS should consider surveying in 2003 a very small, carefully targeted subset of the current panel to study biases in the current sample, possibly to use for the purpose of adjustment.

Recommendation 3: A cost-benefit analysis should be conducted to optimize the relative allocation of resources between the National Survey of College Graduates and the National Survey of Recent College Graduates. Also, additional oversampling should be applied to capture adequate numbers for small domains for which increased interest has become apparent since the last design.

To emphasize the importance of achieving and maintaining high quality of the data, we offer a recommendation on the new sample:

Recommendation 4: The SRS should make every effort to achieve a response rate of 85 percent or higher for the recommended new sample and to retain the sample over time.

We encourage SRS to make every effort to prevent occurrence, in the new sample, of the flaws now present in the old and to conduct methodological research to evaluate the major sources of nonsampling errors, particularly nonresponse error, as well as methods for reducing their effects on the survey estimates. We also offer recommendations pertaining to increasing coverage of populations of interest (e.g., immigrant scientists and engineers and a broader range of fields relevant to science and engineering); refining the definition of goals for SESTAT; regularly monitoring SESTAT data quality; performing more in-house research at SRS; and developing an agenda for subject matter analysis and methodological research.

1

Introduction

This report presents the findings, conclusions, and recommendations of the National Research Council's Committee to Review the 2000 Decade Design of the Scientists and Engineers Statistical Data System (SESTAT). SESTAT is an integrated database resulting from a system of surveys, developed and maintained by the Science Resources Statistics (SRS) division of the National Science Foundation (NSF) to provide timely information about the numbers and characteristics of scientists and engineers in the United States. The committee was convened by the NRC's Committee on National Statistics (CNSTAT) in response to a request by the NSF to review various design options being considered by the SRS staff for SESTAT in the 2000 decade. The committee was charged to conduct a 1-day workshop to bring together SRS, academic, and other experts to examine and discuss the advantages and disadvantages of three analytic designs for SESTAT proposed by SRS. The committee was also charged to conduct a 1-day closed meeting to deliberate on the designs and to produce a brief report of their findings and recommendations.

The committee's information-gathering process included review of *Surveying the Nation's Scientists and Engineers: A Data System for the 1990s* (National Research Council, 1989), as well as careful reading of three key papers sponsored by SRS (Westat 2002a, 2002b, 2002c), which summarize the SRS's deliberations to date on statistical issues pertaining to design options for SESTAT; current and alternate sources of data on the science, engineering, and technical workforce; and comparison of SESTAT esti-

mates of the scientist and engineers population with those from the Current Population Survey. The committee also convened a 1-day workshop with NSF staff and contractors and SESTAT users, representing government and private organizations. The agenda for the workshop is in the appendix to this report.

The committee was not asked, nor did it attempt, to perform an analysis of cost and implementation factors associated with various design options for SESTAT, examine the content of the surveys that support SESTAT, nor revisit the question of how the population of scientists and engineers should be defined.

2

Scope of SESTAT

SESTAT defines scientists and engineers to include both college graduates educated in science and engineering fields and those who work in science and engineering occupations, whether educated in those fields or not. The population of scientists and engineers includes technicians and technologists, researchers, educators, and managers of the science and engineering enterprise (National Science Foundation, 2002; National Research Council, 1989). It totals approximately 12 million individuals, which is about 8 percent of the U.S. labor force. This small group has a disproportional effect on the U.S. economy and particularly on the rapid technological change that characterizes the economy. Consequently, policy makers and researchers within government, industry, and academia need timely information about the numbers and characteristics of scientists and engineers in the United States. Information of interest includes: numbers in each specialty, demographic characteristics and comparisons, numbers entering and leaving different fields, comparison of supply (e.g., numbers of graduates) with demand, matching of education to job, and factors that affect the education and utilization of scientists and engineers (National Science Foundation, 2002; National Research Council, 1989).

In its 1989 report, the National Research Council concluded that NSF should develop and maintain a data system to provide information that permits users to apply their own definitions of the science and engineering population for their particular research and analysis purposes, and that the

system should support analyses of the scientific and engineering community from the perspectives of both occupational employment and academic training (National Research Council, 1989:55-56). NSF describes SESTAT as follows (National Science Foundation, 2002:2):

> SESTAT is a comprehensive and integrated system of information about the employment, educational, and demographic characteristics of scientists and engineers in the United States. It comprises data collected through three national sample surveys supported by NSF: the National Survey of College Graduates (NSCG), the National Survey of Recent College Graduates (NSRCG), and the Survey of Doctorate Recipients (SDR). These surveys are conducted biennially; each is administered to a different sample population of bachelor's and above college degree holders.

THE NATIONAL SURVEY OF COLLEGE GRADUATES

The National Survey of College Graduates (NSCG) was first administered in April 1993 and biennially thereafter, through 1999. The survey covers a nationally representative sample of college degree holders[1] who were identified through the 1990 decennial census. The 1993 NSCG was a special baseline survey of a stratified random sample of individuals identified through the census long form. Eligible persons were those who resided in the United States as of April 1990 and held a bachelor's degree or higher in any field, not necessarily in the sciences or engineering. In 1993, two selected groups from the NSCG were incorporated into the SESTAT database: those with science or engineering (S&E) degrees, and those with non-S&E degrees but who worked in science and engineering occupations during April 1993.

These two populations are collectively referred to as the NSCG S&E panel. In 1995 and subsequent rounds of the survey, these same two groups have been followed (National Science Foundation, 2002; Westat, 2002b).

[1]College degree holders refers to individuals with a bachelor's or higher degree. It does not include individuals with only an associate's degree or specialized degree (e.g., nursing degree below the bachelor's level). In this report the term *S&E degree* refers to bachelor's or higher degree in a science or engineering discipline.

THE NATIONAL SURVEY OF RECENT COLLEGE GRADUATES

A second survey, the National Survey of Recent College Graduates (NSRCG), is used to incorporate into SESTAT new U.S. S&E degree earners in the 1990s. In 1993, the NSRCG consisted of a sample of individuals who earned new S&E bachelor's or master's degrees in the prior three academic years.[2] The NSRCG sample is a cluster sample identified through sampling of educational institutions in a first stage, and, in a second stage, sampling bachelor's degree and master's degree graduates from within these institutions. The Integrated Postsecondary Education Data System (IPEDS) was used to construct the sampling frame for educational institutions. The data from that system, for more than 9,900 postsecondary institutions, includes types of programs, levels of awards offered, enrollments, and degree completions for various levels.

Biennially since 1995, the previous NSRCG sample cases have been moved into the NSCG sample frame, and a new NSRCG sample has been selected, consisting of individuals who earned new S&E bachelor's or master's degrees in the prior 2 academic years.

THE SURVEY OF DOCTORATE RECIPIENTS

A third survey, the Survey of Doctorate Recipients (SDR), initiated during the 1970s, follows a sample of holders of S&E doctorates earned at U.S. institutions throughout their careers, from year of degree award until age 75. Every 2 years, a sample of new S&E doctoral degree earners is added to the SDR from a fourth survey, the Survey of Earned Doctorates (SED). Each SDR sample frame includes all U.S.-earned S&E doctorates through the previous academic year. As new doctorate recipients are added to the panel, the sample of existing participants is dropped to maintain a roughly constant sample size. Table 2-1 presents, for the years 1993 and 1995, the number of cases per survey. Westat (2002b) provides a more detailed description of the surveys that are integrated into the SESTAT, including their sampling techniques and frame sources. Additional detailed information on SESTAT methodology is provided on the SRS SESTAT internet website (http://www.nsf.gov/sbe/srs/stats.htm [September,

[2]The 1993 NSRCG covered those who received college degrees during the period April 1, 1990, to June 30, 1992.

TABLE 2-1 Number of Cases by Survey

Survey	1993		1995	
	Unweighted Cases	Weighted Cases	Unweighted Cases	Weighted Cases
SESTAT Integrated Database	126,721	11,615,200	104,616	12,036,200
NSCG (Full)	148,298	29,021,500	N/A	N/A
NSCG (S&E Panel)	74,462	10,953,100	53,488	10,724,200
NSRCG	19,426	973,400	16,338	841,000
SDR	39,495	513,600	35,370	542,500

SOURCE: National Science Foundation (2002: Appendix B, Table 2)
NOTES: The integrated database is constructed by adding the cases from the NSCG (S&E panel), the NSRCG, and the SDR. The number of cases from the individual surveys does not add to the integrated database total due to overlap of some of the cases.

2002]). Detailed information on IPEDS is available at the website of the Department of Education's Center for Education Statistics (www.ed.gov [September, 2002]).

TYPES OF INFORMATION CONTAINED IN SESTAT

Table 2-2 lists the degree fields and occupational categories within which more detailed educational and occupational labels are organized.

In addition to the estimated counts of individuals for specific educational and occupational categories, SESTAT contains a wide range of information about scientists and engineers (National Science Foundation, 2002). For the employed, labor force information includes items such as primary occupation and salary, type of employer, supervisory responsibilities, relationship between work and highest degree, typical work activities, licensing and certification prerequisites, and secondary jobs held. For those who are unemployed or not in the labor force during the reference week, labor force information includes items such as job last worked, data of last job, and reasons for not working. Other work-related information includes membership in professional associations and participation in work-related training activities.

TABLE 2-2 Degree Fields and Occupational Categories in SESTAT

Major Educational Field and Minor Educational Groups of Disciplines	Major and Minor Occupational Categories
Science and Engineering	
Computer and mathematical sciences • Computer and information sciences • Mathematical sciences	Computer and mathematical scientists • Computer and information scientists • Mathematical scientists • Postsecondary teachers of computer and mathematical sciences
Life and related sciences • Agricultural and food sciences • Biological sciences • Environmental life sciences • Health and related sciences (doctorate only)	Life and related scientists • Agricultural and food scientists • Biological scientists • Environmental life scientists • Postsecondary teachers of life and related sciences
Physical and related sciences • Chemistry, except biochemistry • Earth sciences, geology, and oceanography • Physics and astronomy • Other physical sciences	Physical scientists • Chemists, except biochemists • Earth scientists, geologists, and oceanographers • Physicists and astronomers • Other physical scientists • Postsecondary teachers of physical and related sciences
Social and related sciences • Economics • Political and related sciences • Psychology • Sociology and anthropology • Other social sciences	Social and related scientists • Economists • Political and related scientists • Psychologists • Sociologists and anthropologists • Other social scientists • Postsecondary teachers of social sciences
Engineering • Aerospace and related engineering • Chemical engineering • Civil and architectural engineering • Electrical and related engineering • Industrial engineering • Mechanical engineering • Other engineering	Engineers • Aerospace and related engineers • Chemical engineers • Civil and architectural engineers • Electrical and related engineers • Industrial engineers • Mechanical engineers • Other engineers • Postsecondary teachers of engineering

continued

TABLE 2-2 Continued

Major Educational Field and Minor Educational Groups of Disciplines	Major and Minor Occupational Categories
Nonscience and Nonengineering	
Management and administration	Managers and administrators
Health and related (bachelor's and master's only)	Health-related occupations
Teaching and education	Teachers, except S&E postsecondary teachers
	Non-S&E postsecondary teachers
Social service and related	Social services and related occupations
Technology and technical	Technologists and technicians
Sales and marketing	Sales and marketing occupations
Arts, humanities, and related	Arts, humanities, and related occupations
Other nonscience and nonengineering	Other nonscience and nonengineering occupations

SOURCE National Science Foundation (2002: Appendix A)

Educational information in SESTAT includes dates awarded a high school diploma and associate degrees; level, field, and date for the first bachelor's degree and up to the two most recent other degrees; and any continuing education, including postdegree college courses, field of study, and extent of employer financing. The SESTAT surveys also collect information on family characteristics (marital status, spouse's employment status, children, and parents' educational attainment) and demographic characteristics (age, gender, race/ethnicity, disability, country of birth, and citizenship type).

In addition, special modules have been conducted to provide more detailed information on labor force status, postdoctoral experience, patent and publication activity, and alternative or temporary work experience.

POPULATION COVERAGE OF THE CURRENT SESTAT SAMPLE FRAME

This section describes limitations in the population coverage of the current SESTAT sample, which is based on the 1990 census and supplements from the NSRCG and SDR.

The current SESTAT definition of scientists and engineers includes two disjointed groups: persons with S&E degrees, no matter what their occupation, and college graduates who do not have S&E degrees but who are working in an S&E occupation. Those working in S&E occupations without bachelor's degrees (e.g., persons with associate's degrees in any field) are excluded from the SESTAT population. Figure 2-1 depicts the SESTAT target population. The dashed oval shows the combinations of degree type, labor force status, and S&E occupational status that are included in the SESTAT target population.

Theoretically, the 1993 NSCG can provide complete coverage of the SESTAT population in April 1990, because all members of that population were U.S. residents with college degrees at the time of the 1990 census. However, by 1993, the SESTAT population had grown with the inclusion of four new groups:

(1) Graduates with a first bachelor's degree received in 1990 to 1992, at least one of whose degrees was in an S&E field.
(2) Graduates with a first college degree received in 1990 to 1992, with no college degree in an S&E field, but who were employed in an S&E occupation during April 1993.
(3) Immigrants since April 1990 with a foreign college degree in an S&E field (and no U.S. degree).
(4) Immigrants since April 1990, with one or more foreign college degrees in a non-S&E field, but who were employed in an S&E occupation during April 1993.

The 1993 NSRCG is designed to provide complete coverage of group (1)—recent additions to the population of graduates in an S&E field. However, there is no source available to efficiently sample from groups (2) - (4) over the decade. Consequently, the SESTAT surveys systematically missed those groups.

In subsequent biennial periods, analogous coverage gaps occur. Consequently, the importance of these three omitted groups accumulates over time. In addition, in and after 1995 there was another group that SESTAT systematically failed to capture:

(5) Graduates who were eligible for the 1993 NSCG but not SESTAT (i.e., no S&E degree and not employed in an S&E occupation during April 1993) who became employed in an S&E occupation by April 1995.

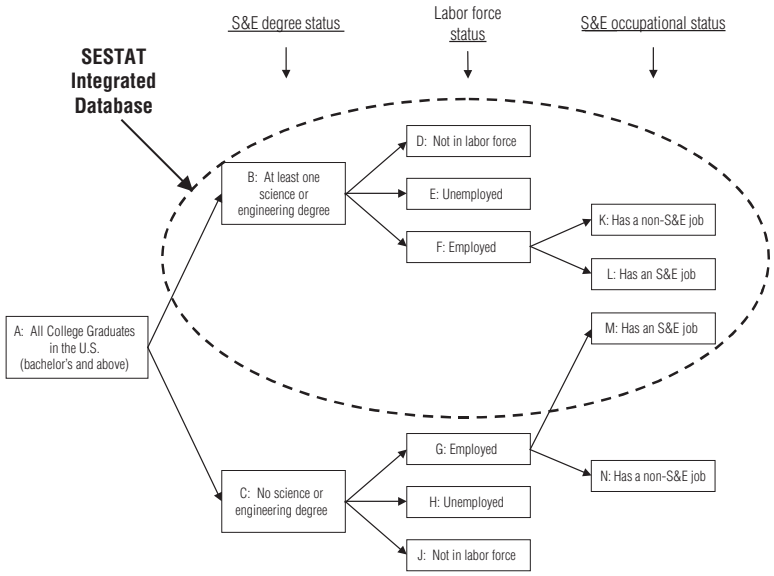

FIGURE 2-1 Schematic of SESTAT population as a function of S&E degree status, labor force status, and S&E occupation status.
SOURCE: National Science Foundation (2002: Figure 1).

Present data do not allow estimation of the size of these noncovered groups with much precision. Almost certainly, group (1)—new graduates with S&E degrees—constituted the majority of additions to the S&E population over the 1990s. However, the noncovered groups may also contribute a substantial portion of the pool. For example, the April 1997 Current Population Survey estimated that 210,000 college graduates working in S&E fields had immigrated into the United States since 1990—6.0 percent of all graduates working in S&E occupations (Westat, 2002c). Although many of those immigrants may have been eligible for the NSRCG, none of those with only foreign degrees would have been covered by SESTAT. Because the majority of the S&E population is not employed in S&E fields, there may be an even larger number of additional post-1990 immigrants with foreign degrees in S&E fields who are not employed in S&E fields.

Survey nonresponse may have an even greater adverse effect than the noncovered groups on the accuracy of estimated characteristics for the S&E population. Table 2-3 summarizes published response rates for the three

SESTAT components. The first column of the table shows response rates for the part of the NSCG sample that SRS attempted to interview. The 78 percent shown for 1993 is the response rate for the initial sample. In subsequent years, SRS contacted only those individuals who had responded in all previous survey waves of the NSCG (or NSRCG, if applicable). Consequently, the NSCG conditional response rates shown for the later years include only respondents from previous cycles. For the last three cycles of conditional rates for NSCG, the unconditional response rates for the initial 1993 NSCG sample (in the second column) are roughly multiplicative.[3] For example, the unconditional response rate in 1995 was about 74 percent (78 percent times 95 percent), and it dropped to about 63 percent in 1999. The response rates shown for the NSRCG and SDR are unconditional, reflecting only the sample that was selected for the particular cycle (year).

The 63 percent unconditional response rate for the NSCG sample in 1999 is troubling. SRS weights data from respondents to account for differences between respondents and nonrespondents in characteristics that are known for both groups. However, to the extent that nonrespondents differ in ways that cannot be explained by known characteristics, SESTAT results will be biased. As the proportion of nonrespondents increased over the 1990s, it is likely that the magnitude of any bias also increased. Unfortunately, SRS does not have data that could help to estimate the size of this bias.

[3]This statement assumes that the conditional response rates for the remaining members of the 1993 NSCG sample are similar to those for new entrants from earlier NSRCG samples.

TABLE 2-3 Selected Unweighted Response Rates for the SESTAT Components (in percent)

	Conditional Response Rate	Unconditional Response Rate		
Year	NSCG (Reported)	NSCG (Computed)[a]	NSRCG (Reported)	SDR (Reported)
1993	78	78	84	87
1995	95	74	83	85[b]
1997	94	70	81	84
1999	91	63	79	82

SOURCE: Data from Westat (2002a: Table 1), amended according to personal communication from SRS staff. Notes: See text for discussion. For background information also see design and methodology report on the National Science Foundation SESTAT website (http://srsstats.sbe.nsf.gov/techinfo.html).

[a]The computed rates for 1995 and later are approximate. The committee acknowledges that these rates are based on assumptions that the committee has not been able to verify.

[b]In 1995 a subsample of mail nonrespondents was selected for computer-assisted telephone interview (CATI) follow-up. The unweighted response rate using the total sample (including those subsampled out) as the base is 77 percent. The weighted response rate was 85 percent, as shown in the table.

3

SESTAT 2000 Decade Design Options

Determining the sample design for the 2003 SESTAT survey marks a critical decision point. The 2000 decennial census offers an opportunity to mitigate the limitations of the current sample by either refreshing or completely replacing the current sample. Another opportunity to do so may not occur for 10 years or longer.

The SRS developed three design options for the 2000 decade (Westat, 2002a); see summary in Table 3-1. The table contains a separate row for each population component of the target SESTAT population. The first eight rows show the population components covered in the current SESTAT sample by one of the past surveys. The next two rows address components of the target population that are currently not covered in SESTAT. The last row addresses scientists and engineers without a college degree**.**

The first column lists subgroups of the general population. The next two columns cover the current SESTAT that falls within these subgroups. The last four columns cover the three design options being considered. The two columns involving SESTAT indicate the frame from which the SESTAT survey was selected. The remaining columns show how each population component would be sampled and surveyed under each of the options.[4]

[4]Westat (2002b, 2002c) identifies a fourth option that focuses on supplementing the old panels. NSF has presented no advantages associated with the fourth option, and the committee has therefore addressed only three options.

TABLE 3-1 Current SESTAT Design and Decade 2000 SESTAT Design Options

	Current SESTAT Design		Decade 2
Population Group	Original Survey	Current Frame	2000 Ce Frame O
Pre-1990 bachelor's and master's	NSCG 1993	1990 postcensal followup	
1991-1992 bachelor's and master's	NSRCG 1993	IPEDS- based	
1993-1994 bachelor's and master's	NSRCG 1995	IPEDS- based	2000 posto
1995-1996 bachelor's and master's	NSRCG 1997	IPEDS- based	
1997-1998 bachelor's and master's	NSRCG 1999	IPEDS- based	
1999-2000 bachelor's and master's	NSRCG 2001	IPEDS- based	2000 pos and Apri 2000 par compone
2001-2002 bachelor's and master's	NSRCG 2003	To be determined	Regular
Doctorates	SDR	SED	SED
Post 1990 census foreign degrees	None	None	2000 pos foreign b oversamp
Post 1990 census non-S&E degrees working in S&E	None	None	2000 pos
Degrees lower than bachelors	None	None	Current only (CP

SOURCE: Westat (2002a: Exhibit 3-1).

NOTE: Acronyms: NSCG, National Survey of College Graduates; NSRCG, National Survey of Recent College Graduates; IPEDS, Integrated Postsecondary Education Data

	Decade 2000 SESTAT Design Options			
me	2000 Census Frame Option	1990 Census Frame Option	Hybrid Option Combining Elements of the 2000 Census Frame Option and the 1990 Census Frame Option	
llowup		1990 postcensal subsample	1990 postcensal subsample	
ed		NSRCG panel subsample (1991-1992 graduates)	NSRCG panel subsample (1991-1992 graduates)	
ed	2000 postcensal	NSRCG panel subsample (1993-1994 graduates)	NSRCG panel subsample (1993-1994 graduates)	2000 postcensal half sample
ed		NSRCG panel subsample (1995-1996 graduates)	NSRCG panel subsample (1995-1996 graduates)	
ed		NSRCG panel subsample (1997-1998 graduates)	NSRCG panel subsample (1997-1998 graduates)	
ed	2000 postcensal and April-June 2000 panel component	NSRCG panel subsample (1999-2000 graduates)	NSRCG panel subsample (1999-2000 graduates)	2000 post-censal half sample and April-June 2000 panel component
mined	Regular NSRCG	Regular NSRCG	Regular NSRCG	
	SED	SED	SED	
	2000 postcensal foreign bachelors oversample	2000 postcensal foreign bachelors targeted subsample	2000 postcensal foreign bachelors oversample	
	2000 postcensal	2000 postcensal targeted subsample	Optional 2000 postcensal targeted sample	2000 post-censal half sample
	Current data only (CPS)	Current data only (CPS)	Current data only (CPS)	

System; SDR, Survey of Doctorate Recipients; SED, Survey of Earned Doctorates; CPS, Current Population Survey

The three options have some characteristics in common. First, the Survey of Doctorate Recipients would continue its current design and system of sampling and data collection under all of the options. Second, under all three options the sample of scientists and engineers would be updated biennially with new graduates from the NSRCG. Third, the SESTAT survey would be conducted biennially under each of these options.

2000 CENSUS FRAME OPTION: NEW NSCG SURVEY BASED ON 2000 CENSUS

Under the 2000 census frame option, a replication of the 1990s design, the Census Bureau would conduct a postcensal survey in October 2003 based on the 2000 census. This survey would include persons who, in April 2000, had received at least a bachelor's degree, were 72 years of age or younger, were not institutionalized, and were living in the United States or overseas serving in the armed forces. In 2003, this sample would be contacted and interviewed for the National Survey of College Graduates. On the basis of the interview, those persons in the 2003 sample who (1) are college graduates with S&E degrees or (2) have a college degree and are working in S&E occupations would be screened into the SESTAT sample. With this approach, persons with foreign S&E degrees who were in the 2000 census as well as those with non-S&E degrees who are working in S&E occupations during October 2003 would be included (Westat, 2002a).

1990 CENSUS FRAME OPTION: A CONTINUATION OF THE 1990s PANELS

Under the 1990 census frame option, the current sample based on the 1990 census would continue, updating gaps in coverage where feasible. Nonrespondents from the original samples since 1993 would be traced in an attempt to decrease bias due to nonresponse. Targeted samples screened from the 2000 census would be used to update the sample of foreign-trained college graduates and those who work in S&E but do not have an S&E degree.

HYBRID OPTION

The hybrid option combines features of the 2000 census frame option and the 1990 census frame option. Part of the sample would be selected using the 2000 census frame option and the remainder using the 1990 census frame option. Under this option, the 2000 census would be used to draw a sample of college graduates of (nominally) about half the size of that planned under the 2000 census frame option. The subpopulations consisting of foreign-trained college graduates and those with non-S&E degrees who have moved into S&E occupations since April 1993 would be represented by this part of the total sample. The remaining portion of the sample would be derived from the existing SESTAT panel, which would be subsampled, bringing its size to (nominally) about half of the total current sample size (Westat, 2002a).

4

Evaluation of the Design Options

EVALUATION CRITERIA

Like most large data systems, SESTAT serves multiple purposes, and no single design can be ideal in all respects. Consequently, the search for the "best" design must balance tradeoffs among multiple objectives. In our evaluation of the three design options for the 2000 decade SESTAT, the committee considered several criteria, adopted from the committee's collective experience with design and evaluation of surveys. As important as deciding on the criteria to consider is determining which should receive priority. Our three primary evaluation criteria relate to the quality of the data going forward:

1. How well will the design cover the complete population of interest?
2. How well will the design achieve and maintain an adequate response rate?
3. Will sample sizes permit sampling precision that is adequate for the principal uses of the survey data?

Given concerns about potential biases in the current sample, criteria 1 and 2 take the highest priority.

Our other evaluation criteria touch on the usefulness of the future (and past) data:

4. Will the design support longitudinal data analysis (e.g., analysis of career changes by individuals) both before and after 2000 and into the future?
5. Will pre- and post-2000 aggregate results be comparable enough to support analysis of trends over time?
6. Will the design provide information that can be used to better understand (and, perhaps, adjust for) biases in the data for the 1990s?

We consider these criteria of secondary importance because each one really matters only to the extent that the quality of future data is high. Although cost is not one of our explicit evaluation criteria, we recognize that cost will constrain parameters of the final design choice.

COMPARISON OF THE 2000 AND 1990 CENSUS FRAME OPTIONS

Before discussing the merits of the hybrid option, we compare the two basic options. Each—the 2000 census frame option and the 1990 census frame option—possesses certain advantages relative to the other.

By replacing the 1990 sample with one chosen from the 2000 census, the 2000 census frame option immediately removes the systematic coverage gaps in the current sample that have accumulated between 1990 and 2000. Thus, the 2000 census frame option will eliminate whatever bias has developed due to cumulative nonresponse. Indeed, the nonresponse problem for the 1990 census frame option may increase dramatically in 2003 because more than 4 years will have passed since the last survey—making recontact more difficult than in previous surveys.

Although the plan for the 1990 census frame option includes strategies to fill in coverage gaps and to reduce nonresponse bias, we expect those to have limited success. Screening with the 2000 census for recent immigrants who meet the SESTAT criteria (groups (3) and (4) detailed above) may be effective. However, we expect that screening for new college graduates in non-S&E fields who work in S&E occupations (group (2)) and for college graduates in non-S&E fields who converted to S&E occupations after April 1993 (group (5)) would be very expensive.[5] Finally, the

[5]It would be an expensive, inefficient use of resources to sample members of group (2) with a probability comparable to that used in the 1993 NSCG. Naively, that would require

proposed attempts to obtain responses from nonrespondents to the 1993 NSCG are unlikely to produce much yield due to the difficulty of tracking over a 10-year period and converting persons who may have previously refused to participate.

In contrast to these disadvantages, the 1990 census frame option offers certain relative advantages. First, it would increase the scope of potential longitudinal analyses. Unlike the 2000 census frame option, the 1990 census frame option would support longitudinal analyses that span the 1990 and 2000 decades, and it would support analyses covering longer periods within the careers of individual scientists and engineers. Although the 2000 census frame option eliminates any potential for current short-term longitudinal analyses, if it is conducted to obtain a high response rate it should improve the quality of those performed on future data. The committee is unaware of much use of the longitudinal data available in the 1990s, so the temporary loss of that capability may not matter much. Of course, there is the possibility that interest in longitudinal analyses based on SESTAT data may increase in the future.

Second, because the 2000 census frame option draws a completely new sample with substantially different methodology than did the 1999 sample, there could be an artificial blip in estimates going from 1999 to 2003, due more to methodology than to real trends. Although the 1990 census frame option promises to produce more stable estimates from 1999 to 2003, this could be mainly because the 2003 data would share most of the biases of 1999.

Third, the 2000 census frame option is expensive. SRS staff estimate that the 2000 census frame option would involve a screening rate (ratio of initial sample size to the number of SESTAT-eligible respondents) of almost 3 to 1, and that this would be the largest screening rate of the three options (Westat, 2002a). Of course, as noted above, the 1990 census frame option involves at least some screening of people sampled from the 2000 census, and that rate might also be quite high. Although cost is not one of the primary evaluation criteria for this committee, resource constraints can

---

an effort comparable to the 1993 NSCG. One could possibly reduce costs by 50-75 percent by focusing on college graduates born after 1965 (most born before then would have obtained a degree before 1990). Consequently, for the 1990 census frame option, SRS would need to sample at a much lower rate in practice, implying a much smaller sample and therefore much larger weights. It would probably be even more difficult to obtain an adequate sample size for group (5), because SRS could not screen on age.

affect sample sizes and, therefore, sampling variability of the estimates produced from a design. However, even if the 1990 census frame option would make more efficient use of resources, that efficiency does not necessarily translate into a reduction in sampling variability because using any cost savings to increase the size of the main components of the sample (the 1993 NSCG and subsequent NSRCGs) is not an option at this time. Furthermore, it is quite possible that the savings from the avoidance of large-scale screening in the 1990 census frame option would be offset by increased costs of maintaining adequate response rates. In addition, recontacting members of the 1990s panel raises a concern about the informed consent procedure in 1999: it appears that at least some participants were informed that they would not be contacted again.

After weighing the pros and cons of these two options, the committee concludes that the 2000 census frame option is a much stronger design than the 1990 census frame option. The most important of the evaluation criteria are those involving data quality. Because the 2000 census frame option provides a direct and effective solution to the problems of the current sample by selecting fresh participants, it should produce much more accurate data in 2003 and into the future. Given our judgment about data quality, the apparent advantages of the 1990 census frame option seem questionable. Based on all these considerations, we believe that it is better to "cut the cord" and collect the best possible data for 2003 and later years.

COMPARISON OF THE 2000 CENSUS FRAME OPTION AND THE HYBRID OPTION

Because the hybrid option is a mixture of the 2000 census frame option and the 1990 census frame option, almost all of the above discussion is relevant for a comparison of it with the 2000 census frame option. Relative to the 2000 census frame option, the hybrid option shares most of the advantages and disadvantages of the 1990 census frame option—only to a lesser extent, depending on the mix of the 1990 and 2000 census frame options in the hybrid. If that were the whole story, then the 2000 census frame option would again be the clear choice.

However, the hybrid approach should not be dismissed so easily. The availability of data from both census frames offers potential advantages that would be unique to this design. In the end, however, we do not believe that those potential advantages are great enough to overcome the liabilities in continuing the current sample based on the 1990 census frame.

SRS has suggested that one potential advantage of the hybrid option is that it may permit an assessment of the bias in the current SESTAT sample (Westat, 2002a: 3-8):

> If the two samples (current panel and 2000 postcensal sample) produce comparable results, then they can be combined with relatively little loss in efficiency. On the other hand, if the comparisons indicate that estimates from the existing panel are much different than those based on the new sample, then it may be presumed that the older sample is biased. In this case, the new sample can still be used to make unbiased cross-sectional estimates (but with reduced levels of precision). Analysis of the differences might also provide improved nonresponse adjustment methodology that would bring the old panels back into effective use.

Assuming that the two samples could be combined successfully, the hybrid approach would allow SESTAT to capitalize on the cost advantages of the 1990 census frame option. The greater cost efficiency of the current panel would allow for a greater overall sample size than in the 2000 census frame option, if the reductions in the cost of screening are not counterbalanced by additional costs of retention. A larger sample size could translate into greater precision. The hybrid option would produce even greater gains for subpopulations of interest (e.g., underrepresented minority groups or small S&E fields) because those groups could be oversampled from the current panel.

The committee believes that it is a mistake to try to combine results from two designs when one is known to have flaws of uncertain consequence. If the two samples produce substantially different results, it would be unwise to combine the two samples without any adjustment. However, any adjustment would be a tricky and likely ineffectual undertaking because there would be no way to accurately attribute the observed difference among sampling error and biases in the two panels. To the extent that the adjustment consists of calibrating the 1990-panel results to the 2000-panel results, the effective sample size would essentially be the size of the 2000 panel. In other words, the data for the 1990 panel would be almost completely wasted.

Even if the two samples produce very similar results for most outcomes, that finding would not provide complete assurance that there is no bias in the data from the 1990 panel. Consequently, it would still be

improper to place much more confidence in the combined results than one would in results based on data from the 2000 panel alone. In short, combining a reduced panel of acceptable quality with another of dubious quality with only a slim possibility of greater understanding of correctable bias is not a wise decision.

An additional potential benefit of the hybrid option is that it would allow SRS to analyze aspects of specific survey items. For example, in 2003, SRS will ask about race using a new format that allows reporting multiple races. Because data for the current sample were collected using a single-race format, comparison of old responses with those on the 2003 NSCG would provide information about the effect of the change in format. Similarly, by asking 2003 NSCG respondents about dates and types of all degrees and comparing that information with earlier data, SRS can learn something about the reliability of this information.

Although we agree that there are important things to learn about a variety of survey items, we do not believe that the hybrid option is the most appropriate source of data. The Census Bureau is conducting research that will more effectively learn about the effects of changes to the race question. Other methodological questions might be addressed more efficiently by focused reinterview studies or other methods.

5

Conclusions and Recommendations

CHOICE AMONG DESIGN OPTIONS

The committee concludes that the 2000 census frame option is the best choice among the available design options.

Recommendation 1: Almost all of the resources allocated to the SESTAT data collection effort in 2003 should be devoted to drawing a new National Survey of College Graduates from the 2000 census and supplementing this panel with the National Survey of Recent College Graduates.

Despite the committee's reservations about combining data from a 1990 panel and 2000 panel, we conclude that it would be valuable to try to learn more about the nonsampling errors that necessarily creep into the system as the original NSCG sample ages. In particular, information about relative biases for the 1999 and 2003 samples is important for purposes of looking at trends across the two decades. Information about noncoverage and nonresponse biases would also help in the decision-making process for future redesigns.

However, the committee is skeptical that much can be learned about the causes of nonsampling error by simply comparing estimates from the old and new panels for the same time period. Although it is possible to

estimate the difference in the biases of the two designs by conducting surveys from the 1990 and 2003 frames simultaneously, exploring the causes of the biases in either design would be quite difficult. Information that is useful for the reduction of nonsampling errors is best obtained from special evaluation studies such as cognitive laboratory investigations, reinterview surveys, record check studies, and reliability analysis. Therefore, the committee questions the usefulness of continuing the old panel indefinitely as suggested in the hybrid option.

As an alternative to continuing the old panel indefinitely, the committee suggests carrying forward a small sample of the old panel on a one-time basis if the SRS staff considers it important to do so for the following purposes: (1) using comparisons of the new and old panel estimates to generate hypotheses regarding the sources and causes of the nonsampling error that could then be explored in more detail using the special evaluation study methods noted above; and (2) using simple ratios of the estimates derived from the old and new samples in 2003 to adjust estimates from the old data series so that they can be compared with estimates from the new design. In this way, the biases in contrasts of estimates from the old and new data series due to survey design differences could be substantially reduced.

Recommendation 2: If SRS staff confirm that a targeted sample could be useful for the purpose of adjustment, SRS should consider surveying in 2003 a small and carefully designed subsample of the current panel to study biases in the current sample, possibly to use for the purpose of adjustment.

This sample could help direct efforts to evaluate nonsampling error in the new panel and would also provide a means to adjust the old data series for comparisons with the new data series.

Recommendation 3: A cost-benefit analysis should be conducted to optimize the relative allocation of resources between the National Survey of College Graduates and the National Survey of Recent College Graduates. Also, additional oversampling should be applied to capture adequate numbers for small domains for which increased interest has become apparent since the last design.

EMPHASIS ON DATA QUALITY

The committee urges emphasis on the quality of the 2000 census panel because there is no guarantee that the 2010 census will have a long form, nor that the American Community Survey will be implemented in a form that adequately supports the SESTAT database. Every effort should be made to prevent occurrence, in the new sample, of the flaws now present in the old sample. For example, it is possible that fast developing technology such as Internet communication can be used in an innovative way to provide feedback to respondents and greater incentive for them to remain in the survey. Of course, this and other enhanced follow-up efforts may increase the per-case costs of data collection.

The survey should be designed to achieve the highest response rate possible subject to cost, timing, and data quality constraints. There are no absolute standards for response rates in the survey community. However, given that SESTAT involves surveys of a well-educated population about topics that are salient to them, an 85 percent (weighted) response rate is a reasonable goal and consistent with response rates routinely obtained by the Census Bureau in other surveys. One difficulty will be finding persons enumerated in the 2000 Census after 3 or more years. Assuming a 90 percent response rate for those census persons who can be located for the survey (i.e., 90 percent combined contact/cooperation rate for locatable persons), the rate at which persons can be located would have to be around 95 percent to achieve an overall survey response rate of 85 percent. We believe this fairly ambitious goal is achievable.

Recommendation 4: The SRS should make every effort to achieve a response rate of 85 percent or better for the recommended new sample and to retain the sample over time.

Toward that end, the SRS should consider trading off sample size and, accordingly, accepting some additional degree of sampling variability to improve the quality of the data collected if promising concrete steps in that direction can be identified.

Recommendation 5: SRS should conduct methodological research to evaluate the major sources of nonsampling errors, particularly nonresponse error, as well as methods for reducing their effects on the survey estimates.

COVERAGE OF RELEVANT POPULATIONS

Although it is beyond the committee's scope to review the appropriateness of SESTAT's definition of the population of scientists and engineers, we recognize the valuable contributions of others working in fields that touch upon science and engineering, and that more can be learned about scientists and engineers if information is available to permit comparisons against the population of workers as a whole.

Recommendation 6: SRS should pursue a plan to carry forward non-S&E individuals in the panel, including the fields of health, S&E education, and possibly other non-S&E fields that relate to S&E. The committee encourages efforts (including the seeking of funds) to also include non-S&E education and, if possible, other non-S&E fields.

The committee agrees with SRS that immigrants are an important, undercovered, and increasing part of the S&E community, so it is important to try to include recent immigrants with foreign degrees in the sample.

Recommendation 7: SRS should investigate productive means of reaching the goal of greater inclusion of immigrants in the sample, including, if feasible and productive, cooperative work with the Immigration and Naturalization Service.

PLANNING EXPANDED SRS ACTIVITIES AND GREATER INTERACTION WITH COLLEAGUES

The committee commends the SRS for completing the weights pertaining to longitudinal data.

Recommendation 8: SRS should encourage utilization of the longitudinal nature of the SESTAT data.

Looking ahead, the committee believes that taking certain steps now will help the SRS to plan for future SESTAT surveys over the 2000 decade and beyond.

Recommendation 9: SRS should prepare a concise, clear, and complete statement of the goals for the SESTAT database. The statement of

goals should reflect the SRS's understanding of priorities that take into account the conflicting interests that compete for resources related to the collection of the SESTAT data. SRS goals should include expanding and maintaining meaningful contact with academic colleagues and with users from all sectors, including academia, business, and government.

Recommendation 10: SRS should regularly monitor the quality of the SESTAT data and perform more in-house research to foster increased intellectual curiosity, deep knowledge of the SESTAT data, and data-driven re-design activities. To this end, SRS should develop a well-planned agenda for subject matter analysis and for methodological research.

References

National Science Foundation (2002). *SESTAT: A Tool for Studying Scientists and Engineers in the United States.* Available electronically: www.nsf.gov/sbe/srs/nsf99337/secta.htm [September, 2002].

National Research Council (1989). *Surveying the Nation's Scientists and Engineers: A Data System for the 1990s.* Panel to Study the NSF Scientific and Technical Personnel Data System, Committee on National Statistics. Washington, DC: National Academy Press.

Westat (2002a). Design Options for SESTAT for the Current Decade: Statistical Issues. Draft report to the Science Resources Statistics Division, U.S. National Science Foundation, Washington, DC.

Westat (2002b). Current and Alternate Sources of Data on the Science, Technical, and Engineering Workforce. Draft report to the Science Resources Statistics Division, U.S. National Science Foundation, Washington, DC.

Westat (2002c). Comparison of the Scientists and Engineers Statistical Data System with the Current Population Survey. Draft report to the Science Resources Statistics Division, U.S. National Science Foundation, Washington, DC.

Appendix

AGENDA

SESTAT Workshop: July 9, 2002

9:00 Introduction

Andrew White, Director, Committee on National Statistics, The National Academies
Norman Bradburn, Director for the Social, Behavioral, and Economic Sciences Directorate, National Science Foundation
Lynda Carlson, Director, Division of Science Resources Statistics, National Science Foundation
Robert Bell, Chair, Committee to Review the 2000 Decade Design of the SESTAT

9:20 SESTAT Background

9:20 1990s surveys with emphasis on survey process: *Ron Fecso,* Division of Science Resources Statistics, NSF
- Committee Q&A

9:50 Summary of 1989 NRC panel review and expectations—*Graham Kalton*, Westat
- Committee Q&A

10:20 Break

10:30 SESTAT Background (continued)

- 10:30 Overview of Redesign Activities for 2000 Decade, Including Content—*Lynda Carlson,* Director, Division of Science Resources Statistics, National Science Foundation
  Mary Frase, Deputy Director, Division of Science Resources Statistics, National Science Foundation
  - Committee Q&A

11:00 Sample Design for 2000 Decade

- 11:00 Frames and Coverage Issues—*Nirmala Kannankutty,* Division of Science Resources Statistics, National Science Foundation
  - Committee Q&A

12:00 Lunch–Committee Caucus

1:00 Sample Design for 2000 Decade (continued)

- 1:00 Design Options—*Ron Fecso,* Division of Science Resources Statistics, National Science Foundation
  - Committee Q&A

2:30 Break and Committee Caucus

2:45 Pros and Cons of Design Options: Committee Interacts with Presenters

4:10 Audience Q&A and Comments on Design Options

4:30 Adjourn