



A Strategy for Assessing Science: Behavioral and Social Research on Aging

Committee on Assessing Behavioral and Social Science Research on Aging, Irwin Feller and Paul C. Stern, editors, National Research Council

ISBN: 0-309-66759-3, 176 pages, 6 x 9, (2006)

This PDF is available from the National Academies Press at:
<http://www.nap.edu/catalog/11788.html>

Visit the [National Academies Press](http://www.nap.edu) online, the authoritative source for all books from the [National Academy of Sciences](http://www.nap.edu), the [National Academy of Engineering](http://www.nap.edu), the [Institute of Medicine](http://www.nap.edu), and the [National Research Council](http://www.nap.edu):

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Explore our innovative research tools – try the “[Research Dashboard](#)” now!
- [Sign up](#) to be notified when new books are published
- Purchase printed books and selected PDF files

Thank you for downloading this PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, [visit us online](#), or send an email to feedback@nap.edu.

This book plus thousands more are available at <http://www.nap.edu>.

Copyright © National Academy of Sciences. All rights reserved.

Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. [Request reprint permission for this book.](#)

A STRATEGY FOR ASSESSING SCIENCE

Behavioral and Social Research on Aging

Committee on Assessing Behavioral and Social Science Research on Aging
Irwin Feller and Paul C. Stern, Editors

Center for Studies of Behavior and Development
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, N.W. 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 project was supported by Award No. NO1-OD-4-2139, Task Order No. 122 between the National Academy of Sciences and the Department of Health and Human Services. 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.

Library of Congress Cataloging-in-Publication Data

A Strategy for Assessing science : behavioral and social research on aging / Committee on Assessing Behavioral and Social Science Research on Aging, Center for Studies of Behavior and Development, Division of Behavioral and Social Sciences and Education ; Irwin Feller and Paul C. Stern, editors.

p. ; cm.

Includes bibliographical references.

ISBN-13: 978-0-309-10397-8 (pbk.)

ISBN-10: 0-309-66759-3 (pdf)

1. Gerontology—United States. 2. Aging—Government policy—United States. I. Feller, Irwin. II. Stern, Paul C., 1944- III. Center for Studies of Behavior and Development (U.S.). Committee on Assessing Behavioral and Social Science Research on Aging.

[DNLM: 1. Aging—United States. 2. Research Design—United States. 3. Behavioral Research—United States. 4. Cognition Disorders—United States. WT 20 A846 2007] HQ1064.U5A87 2007

305.260973—dc22

2006039253

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, 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.

Copyright 2007 by the National Academy of Sciences. All rights reserved.

Suggested citation: National Research Council. (2007). *A Strategy for Assessing Science: Behavioral and Social Research on Aging*, Committee on Assessing Behavioral and Social Science Research on Aging. Irwin Feller and Paul C. Stern, Editors. Center for Studies of Behavior and Development, 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. Ralph J. Cicerone 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. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON ASSESSING BEHAVIORAL AND SOCIAL
SCIENCE RESEARCH ON AGING

IRWIN FELLER (*Chair*), American Association for the Advancement of
Science, Washington, DC

WENDY BALDWIN, Office of the Vice President for Research, University
of Kentucky

PAUL B. BALTES, Max Planck Institute for Human Development, Berlin

RICHARD DeVEAUX, Department of Mathematics and Statistics,
Williams College

JAMES JACKSON, Department of Psychology, University of Michigan,
Ann Arbor

JANICE KIECOLT-GLASER, Department of Psychiatry, Ohio State
University

ROBERT E. KOHLER, Department of History and Sociology of Science,
University of Pennsylvania

MICHELE LAMONT, Department of Sociology, Harvard University

LEAH L. LIGHT, Department of Psychology, Pitzer College

DANIEL McFADDEN, Department of Economics, University of
California, Berkeley

GARY SANDEFUR, College of Letters and Science, University of
Wisconsin, Madison

SHRIPAD TULJAPURKAR, Department of Biological Sciences, Stanford
University

GEORGE E. WALKER, Carnegie Foundation for the Advancement of
Teaching, Stanford, CA

CAROL WEISS, Graduate School of Education, Harvard University

DAVID WISE, John F. Kennedy School of Government, Harvard University

PAUL C. STERN, *Study Director*

LINDA A. DePUGH, *Administrative Assistant*

Dedication

This report is dedicated to the memory of Paul B. Baltes, director of the Center of Lifespan Psychology at the Max Planck Institute for Human Development, Berlin, who died on November 7, 2006.

Paul was an active contributor in the committee's early meetings, before illness limited his further participation in person. His expertise in multiple facets of aging research, breadth of perspective on the science policy and organizational issues embedded in the committee's charge, and above all, skill, warmth, and civility in helping forge common ground on which individuals from disparate fields could base their analysis and recommendations were singular contributions that suffused subsequent committee meetings and the preparation of this report.

Preface

This report responds to a request from the Office of Behavioral and Social Research (BSR) at the National Institute on Aging for a study on how best to assess the progress and vitality of areas of behavioral and social science research on aging and on how to identify the factors that contribute to the likelihood of discoveries in areas of aging research.

BSR's request embodies both some of the longest standing and most current of questions confronted in the formulation of national science policies, in both the United States and other countries. The request includes criteria questions, such as what kinds of science should the public sector, or specific agencies, fund; selection mechanism questions, such as what procedures should be used to implement these criteria; principal-agent decision questions, such as which individual(s) or groups of individuals should possess the authority to make decisions regarding choice of areas of funding or selection of specific research proposals; conditions for success questions, such as the size and composition of the most productive research unit, ranging from single investigators to large teams comprised of researchers to several disciplines; and quality assessment questions, such as the compatibility between established disciplinary-based procedures for organizing selection panels and assessing the importance of scientific findings with statements about the increased salience of research done at the intersections of fields or the interstices between and among them.

These questions in part derive from broad trends in the United States and elsewhere toward increased demands for accountability and documentation of performance on the part of government agencies across all functional areas, including public-sector support of science and technology. In

this respect, the above questions connect logically to those subsumed within the Government Performance and Results Act, the President's R&D Investment Criteria, and the Office of Management and Budget's Performance Assessment Rating Tool (PART).

There is a special salience to BSR's request. From the perspective of a single federal agency and program officer, it poses many of the very same questions that are latent in recent calls by John Marburger, director of the Office of Science and Technology Policy, for a "new science of science policy." Addressing the specific operational needs of a single office in the context of the larger historical, policy, and analytical discourse on criteria and mechanisms for setting research priorities, evaluating returns from past investments and identifying the factors that contribute to research productivity is obviously not simple. As noted above, BSR's request connects to long-standing, generic issues of science policy. The committee's task (and obligation) was to be responsive to the specific charge from a specific sponsor.

The report has attempted to address both the general and the specific aspects of BSR's request. It places the request within the larger and long-standing search for criteria and methodologies for assessing the vitality and performance of fields of scientific inquiry and determining the conditions that lead to scientific success. At the same time, it addresses BSR's mission to support behavioral and social science research on aging, the organizational context in which it operates, and the fields of research it supports. Likewise, the report's recommendations are directed specifically at meeting BSR's programmatic concerns. Retracing at selected points well-known issues, the report also makes more explicit than earlier studies and recent reports several of the organizational, political, and methodological issues that permeate and beset debates about criteria for scientific choice. From this vantage point, it notes how its findings and recommendations connect to larger science policy themes, including areas of needed research to strengthen the scientific basis on which science policy decisions are made.

Attending to both the specific and the general intellectual and policy richness embedded in BSR's request for this study inevitably requires trade-offs about breadth and depth of coverage of selected topics. The committee's choices and the rationales behind them are detailed in the body of the report. In general, to increase the near-term prospective usefulness of its recommendations, the committee has chosen to focus its review of methodological techniques on those now used or considered by U.S. federal science agencies. Coverage is provided of a larger range of research forecasting and assessment techniques, such as used by U.S. industry and by European countries, but for reasons noted not with extensive detail.

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 National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its 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 review of this report: James Banks, Professor of Economics, University College London and Institute for Fiscal Studies; Don Brenneis, Department of Anthropology, University of California, Santa Cruz; Margaret Gatz, University of Southern California; Robert Hauser, Center for Demography, University of Wisconsin–Madison; Diana Hicks, School of Public Policy, Georgia Institute of Technology; Guohua Li, Department of Emergency Medicine, Johns Hopkins University; Duncan T. Moore, Institute of Optics, University of Rochester; Zur Shapira, Stern School of Business, New York University; and Neil Smelser, Department of Sociology, University of California.

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 Marshall S. Smith, Education Program, The William and Flora Hewlett Foundation. 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.

Irwin Feller, *Chair*
Committee on Assessing Behavioral and
Social Science Research on Aging

Contents

Executive Summary	1
1 The Purpose of the Study	7
The Committee's Charge, 9	
Techniques and Processes for Science Assessment, 11	
Goals of the Study, 14	
Organization of the Report, 17	
Notes, 18	
2 The NIA Behavioral and Social Research Program	20
Strategic Goals of NIA, 20	
The Diverse BSR Research Portfolio, 22	
Decision-Making Processes at BSR, 28	
Research Review in NIH, 33	
Notes, 38	
3 The Stakes in Research Assessment	40
Brief History of Federal Science Priority Setting, 42	
Debate over Priority Setting and Assessment Mechanisms, 47	
Research Assessment and the Issue of Power, 59	
Notes, 65	

4	Progress in Science	67
	Theories of Scientific Progress, 67	
	Nature of Scientific Progress, 70	
	Indicators of Scientific Progress, 80	
	Factors That Contribute to Scientific Discoveries, 84	
	Implications for Decision Making, 89	
	Notes, 92	
5	Methods of Assessing Science	95
	A Framework: Analysis and Deliberation as Assessment Strategies, 97	
	Analytical Methods, 99	
	Deliberative Methods, 108	
	Analytic-Deliberative Methods, 111	
	Conclusions and Research Needs, 116	
	Notes, 123	
6	Conclusions and Recommendations	124
	Conclusions, 126	
	Principles for Priority Setting, 130	
	Recommendations, 132	
	References	137
	Appendix	
	Biographical Sketches of Committee Members and Staff	157

Executive Summary

Priority setting is a difficult, perennial issue in science policy, made more difficult in times of tightening research budgets. This report responds to a request from the Office of Behavioral and Social Research (BSR) at the National Institute on Aging (NIA) for advice on how best to judge the rates of progress in the research fields the office supports and to evaluate whether and how to shift the balance of research investments across fields. The request partly reflects a concern that traditional expert review processes are too strongly influenced by established disciplines and fields and too conservative in relation to the need to support research that might generate scientific breakthroughs.

In developing our recommendations, we considered available knowledge about how science makes progress, which shows great variety in types of progress and paths to progress, as well as the considerable difficulty of accurately anticipating these paths. Research areas that appear at one time to be “hot” may prove in retrospect to have been fads, and fields that appear unproductive may be stagnant, fallow, or pregnant. Accurate foresight is very difficult to achieve. We considered decision-making strategies that could address the sponsor’s concerns, along with other legitimate science policy concerns about the quality and rationality of the decision process, the accountability of decision making, and the appropriate balance of influence between scientific communities and agency science managers. Our recommendations are addressed to BSR, but we think the decision strategy we propose is appropriate for a wider range of federal science agencies within and beyond the National Institutes of Health (NIH).

Two generic decision strategies are available for assessing scientific progress and setting research priorities: (1) applying analytic techniques, such as benefit-cost analysis, bibliometric analysis, and decision analysis, and (2) using deliberative processes, such as those of expert review panels. Many methods, both analytic and deliberative, can have value for assessment and decision making, but they all also have limitations. Quantitative analytical methods typically have limitations associated with data collection, reliability, validity, cost, timeliness, and acceptability, as well as the lack of knowledge about how best to combine measures of qualitatively different aspects of scientific progress. Qualitative methods of deliberation are of unknown reliability and may be highly dependent on who is involved in the deliberations and how the questions for deliberation are framed.

Because of the uncertainties about the reliability and validity of all existing methods, we recommend a strategy for decision making that relies primarily on processes and secondarily on methods. It uses techniques based on decision research to structure and inform deliberation within groups of scientific advisers and agency decision makers and to make communication between such groups more transparent, for example, by clarifying the sources of any disagreements in judgment between them. Analytic techniques for quantifying scientific progress can provide useful input to decision-making deliberations, but they should not be used as substitutes for the necessary judgment.

We recommend a strategy that combines analysis and deliberation, in which processes of open, explicit dialogue are organized to raise all the major decision-relevant issues, allow for input from all relevant perspectives, and provide for iterative discussion between researchers and science managers and for orderly reconsideration of past decisions. Such dialogue can also provide for improved accountability of decision making.

Three principles should guide BSR practice in setting priorities across research fields:

Explicitness. *Judgments about the progress and potential of scientific fields should be based on explicit consideration of them in relation to all the major scientific and societal goals of the BSR Program and all the major processes and inputs supporting progress in each field.*

Perspective. *Both extramural research scientists and institute program managers should be involved in assessing the progress and potential of the research fields supported by the BSR Program. Both sets of contributors to priority-setting decisions bring valuable knowledge and insights to the process, as well as different, complementary perspectives.*

Iteration. *Priority-setting exercises should be conducted regularly, and they should include reconsideration of past decisions.*

We make the following specific recommendations for implementing these principles:

1. The staff of the BSR Program, with the help of the program's scientific advisers, should develop an explicit list of scientific outcome and societal impact goals for the program in line with the strategic program goals of NIA. Information from the staff to advisory groups regarding the progress of program-supported research should reference these goals.

2. NIA should periodically conduct a general assessment of the BSR Program with respect to its overall adequacy for supporting the program's scientific outcome and societal impact goals.

Assessments should be conducted approximately every four years and should consider each program goal in relation to each aspect of the BSR Program judged to be important for achieving it (e.g., the different kinds of research activities supported and modes of support).

3. NIA should periodically conduct an area-based assessment of the BSR Program that includes recommended priorities for new and continued support among the substantive areas of research included in the program. These efforts should explicitly assess and compare the past and potential contributions of research in each area receiving major BSR support with regard to each of BSR's goals for scientific outcome and societal impact and with respect to the various inputs and processes that contribute to achieving the goals.

These assessments should also be conducted approximately every four years. They should make recommendations as appropriate for each area on issues of portfolio allocation between disciplinary and interdisciplinary research; basic and applied research; high-risk and low-risk research; development of research methods, of data, and of findings; support of research centers, program projects, and individual investigators; and support of research, infrastructure, and human resources development.

4. The BSR program director should consider the area-based assessments and recommendations carefully in reallocating funds among fields. One year after completion of each area-based assessment, BSR staff should report on decisions reached and actions taken that involve priority setting among research areas and portfolio allocation within areas. The report should explicitly discuss the justification for program decisions that might seem inconsistent with the assessment's recommendations. The report should be delivered to the NIA director and the NIA advisory council.

There can be good justifications for institute decisions that deviate from the recommendations of a body of scientists. The purpose of the recommended report is to ensure that such justifications are made explicit and thus to provide increased accountability in an institutional sense and a continuing rational dialogue among scientists and program managers, focused on the program's objectives.

5. The NIA BSR Program, together with the rest of NIA and NIH, as well as the National Science Foundation and other federal science agencies, should support a coordinated program of research to promote well-informed, high-quality research policy making.

This research program would provide knowledge of broad value to federal science policy and contribute to development of what has been called a "social science of science policy." It is for this reason that we recommend that a broad range of federal science agencies support this research program. The research should pursue three objectives:

a. *Improve basic understanding of scientific progress and the roles of research funding agencies in promoting it.* Research pursuing this objective would examine the nature and paths of progress in science, including the roles of decisions by science agencies. It might include historical analyses of the evolution of scientific fields; advanced bibliometric analyses of the development of research fields over time and the flows of influence among them; studies of the effects of the structure of research fields on their progress; studies of the roles of officials in science agencies in scientific progress; studies of how expert advisory groups, including study sections and advisory councils, make decisions affecting scientific progress; and studies of the effects of the organization of such groups on their success at promoting interdisciplinary and problem-focused scientific activity and ultimately at improving scientific outcomes and societal impacts. In the case of BSR, the research should focus on progress in fields of behavioral and social science related to aging.

b. *Improve understanding of the uses of quantitative decision aids in making research policy decisions.* This research should include the development, trial use, and empirical investigation of the use of quantitative measures and decision-analytic techniques as inputs to priority setting. It should not seek techniques that can supplant deliberation, because different areas of science make different kinds of progress and judgment will always be required to assess progress against multiple objectives. The research would aim to identify useful techniques and determine how to use them effectively. The research might include studies to assess the value of providing information developed through specific analytic techniques; studies comparing

indicators of scientific progress with each other and with unaided expert judgment; comparative quantitative studies of fields that are widely judged to differ in rates of progress; tests of ways to combine information from different analytic methods; and studies of the use of qualitative decision analytic techniques for guiding deliberation.

c. *Develop useful techniques for systematic deliberation in advisory and decision-making procedures.* This research would explore and assess techniques for structured deliberation, some of them including the use of indicators of scientific progress and potential, for retrospective assessment and for priority setting. It would be used to elaborate and refine deliberative methods now in use and those recommended in this study. It should include studies that apply techniques for structuring deliberation to the research priority-setting tasks facing BSR; studies of trials in which review and advisory panels are instructed or trained to focus their deliberations on how each research field might contribute to specified program objectives or goals, including both those related to scientific quality and to mission relevance; studies of attempts to adapt the NIH Consensus Development Conference model to research priority setting; studies comparing advisory panels of different composition; and studies of the effects of instruction and training of advisory panel members to consider the full range of BSR and NIA objectives.

1

The Purpose of the Study

The U.S. federal government supports scientific and technological research to address a broad range of national needs and objectives and to gain fundamental understanding of the processes that shape the world in which people live. Each federal science agency promotes scientific progress toward these objectives in the areas of its mission responsibilities. This is done most obviously by providing funds for research and its infrastructure, including the education and training of succeeding cohorts of researchers; by organizing and setting rules for the external groups that advise on worthy research investments; and by setting research priorities and making choices among specific research programs and projects. Each agency also does so by recruiting, training, and evaluating research managers for their scientific expertise and managerial skills. Agencies redirect support among lines of research when opportunities arise to open new and exciting paths to knowledge and societal benefit, when changes occur in the relative importance of the results from past investments in research, and when specific lines of inquiry or modes of research support are deemed no longer to be productive.

Historically, federal government support has been instrumental in the development of important new fields of science and technology, such as materials science and computer science. Less well understood is how closely the rise or demise of a research field may be tied to federal support. At any point in time, numerous research fields are in an embryonic state and are potential candidates for further maturation. Not all flourish, however, beyond the involvement of small cadres of researchers subsumed within larger subspecialties and disciplines. Support from federal government agencies can

make the difference between development or stagnation for embryonic or fledgling research fields. Priority-setting decisions by federal science agencies thus affect the vitality of existing fields of research, although the strength of this effect is not well known.

Because society has limited resources to support scientific activities, assessing scientific progress and setting priorities are perennial practical components of national science policy decision making. Perennial questions arise, too, about matching agency funding practices with the conditions perceived to be most likely to lead to program or project success in terms of contributions to scientific knowledge and societal objectives. Which objectives most deserve support for science—defense, economic competitiveness, energy, health, the environment, or some other? How should funds be distributed across agencies? Which fields of science most require or merit public-sector support? And which modes of research support are most productive? These questions raise issues of outcomes—assessing the likely benefits¹ of the scientific work being supported—and processes—ensuring that decisions are made in ways that satisfy the criteria of equity, transparency, expertise, and accountability normally demanded of public decisions in a democratic society.

Concerns about the processes and outcomes of priority setting in U.S. government research agencies have intensified in recent years because of a confluence of several clusters of influences. One of these relates to increased demands for accountability and documented performance throughout government, extending across all functional activities, including support of research. For federal government agencies, these demands are most visible in the requirements of the Government Performance and Results Act (GPRA) of 1993 and the Office of Management and Budget's Performance Assessment Rating Tool (PART) process. Demands for accountability, amounting in some accounts to an audit explosion (Power, 1997), have also become more acute because of a recent deceleration or reversal of research budget growth across many functional areas of federal government activities, except for defense and homeland security. In the specific case of the National Institutes of Health (NIH), it is evident that the recent era, which produced a doubling of funds, has ended. The success rates of research grant applications at NIH have declined rapidly over the past several years: for new applications, from around 20 percent in 1999-2002 to 9 percent in 2005; for renewal applications, from around 50 percent to 32.4 percent in the same period (Mandel and Vesell, 2006).

An important source of the renewed interest in reexamination of agency priority-setting criteria and processes has been the endeavors of federal agencies and their division directors and other managers for continuous improvement of the quality of their programs. In their roles as professional science managers, agency officials seek to direct agency resources to support

missions most effectively and efficiently. In doing this, they must simultaneously work to satisfy the priorities expressed by the administration and Congress via the budget process, adhere to the related legislative and administrative requirements on the allocation and expenditure of federal research support, respond to the communities of researchers currently active in relevant fields of inquiry, and in periods of scientific transformation attract and nurture researchers whose work connects to their agency's mission—all this while being responsive to often unpredictable changes in the potential for progress along different paths of scientific inquiry. Science managers do this differently depending on the responsibilities, powers, and activities associated with their positions in their agencies (Seidman, 1998).²

Related to both these influences are increased demands for evidence-based decision making and particularly for decision making based on quantitative evidence. These demands are contained in the provisions of GPRA and PART. They also appear implicitly in the call by the president's science adviser and the director of the Office of Science and Technology Policy, John Marburger (2005), for a "social science of science policy" that would, among other things, use econometrics and other social science methods to help examine the effectiveness of federal investments in science (American Sociological Association, 2006).

Yet another stimulus for renewed attention to priority setting has been the beliefs, latent in recent assessments of the state of U.S. science, that the decision processes of science agencies are unduly conservative in program and project selection and that they fall short in converting research findings into usable and useful applications. When coupled, these two propositions imply that the national investment in scientific research is not yielding its expected returns in improvements in the quality of life and suggest that the United States may lose its preeminence in world science.

THE COMMITTEE'S CHARGE

This national context sets the framework for this report, which responds to the specific needs of the Behavioral and Social Research Program (BSR) of the National Institute on Aging (NIA) to assess the progress and prospects of behavioral and social science research on the processes of aging at both the individual and societal levels. Specifically, BSR asked the National Academies to organize this study with two major goals: "to explore methodologies for assessing the progress and vitality of areas of behavioral and social science research on aging . . . [and] to identify the factors that contribute to the likelihood of discoveries in areas of aging research" (National Academies' proposal to NIA, 2003).

Contained within these two major goals are several specific questions and subthemes, including the following: Given increasing pressures for ac-

countability and for research to have broader impacts, widening choices of research areas to support, and the resulting increased competition for funds, what information can research managers rely upon to guide their allocations of research resources? How can they more effectively advance scientific disciplines and other research areas and make important discoveries more likely? Can we measure or at least compare the progress in different disciplines and research areas? Can the vitality of research areas be defined and assessed? What indicators for fields, as well as for individuals, would be useful? Can progress be effectively tracked through discoveries? How would discoveries be determined and selected for this purpose? BSR has requested advice on methods for the retrospective assessment of scientific progress and for addressing the prospective problem of priority setting to promote the future progress of areas of research.³

We have addressed the two major issues for this study by reviewing and assessing relevant literatures and techniques. We also commissioned a special pilot project to assess the validity and feasibility of newly developed bibliometric methods for assessing research progress.

This study is being conducted at a time of considerable ferment and disagreement about the optimal portfolio of research funding mechanisms. In a stylized (and at times unduly polarized) manner, the choice is presented as between single investigator-initiated, discipline-based proposals and multidisciplinary, team-based proposals prepared in response to an agency request, referred to occasionally as Mode I and Mode II forms of research (Gibbons et al., 1994). As detailed in the following chapters, endorsements of the merits of each approach (and implicit or explicit criticism of other funding approaches), as well as advocacy of numerous intermediate arrangements, are easy to find in current statements on science policy. Systematic empirical work that would permit more evidence-based assessments of policy options, however, is not easy to find, either generally across areas of federally funded research or specifically with respect to aging research.

Both conceptually and empirically, the two components of our charge overlap. For example, to determine the factors that contribute to scientific discoveries, one must rely on the same set of methods (e.g., peer review, bibliometrics) that are used for assessing the progress and potential of these fields and for informing research policy decisions about them. We have therefore collapsed much of the discussion of both elements of the charge into the treatment of assessment methodologies, while separately discussing bodies of knowledge specifically addressed to factors deemed to contribute to scientific advances, especially as they may pertain to research on aging. As those discussions make clear, there are significant gaps and uncertainties in knowledge about the factors that contribute to scientific advances, so that considerable interpretation and judgment are necessary in evaluating the past progress of science or projecting future prospects.

TECHNIQUES AND PROCESSES FOR SCIENCE ASSESSMENT

Interest in methods for setting science priorities on the part of Congress, executive branch units, and agency science program managers has a long history (early contributions include Scherer, 1965; Nelson, 1959; National Academy of Sciences 1965; Rettig et al., 1974; U.S. Congress, Office of Technology Assessment, 1986; a more recent effort is National Science and Technology Council, 1996). Interest continues because of the continued salience of the underlying questions and a widespread belief that, for all their sophistication, the existing studies do not fully satisfy the needs of science program managers for reliable and defensible methods for making priority-setting decisions.

Interest in identifying the conditions that lead to advances in the social sciences also has a long, distinguished pedigree. Antecedents may be found in the 1933 President's Research Committee on Social Trends (see Gerstein, 1986; Smelser, 1986). A more recent inquiry along these lines was the work of Deutsch et al. (1971), who identified and analyzed the conditions that underlay 62 "major advances in the social sciences." This work stimulated a continuing line of inquiry that has sought to distill the relative influence on the conduct of research of such factors as whether the research was conducted by individuals or teams; whether it focused on theory, method, or empirical study; the age of the researcher; and the use of capital equipment. This line of research has confronted but not resolved important methodological and conceptual issues, such as how to select "advances" for study and distinctions between "discovery" and "application" (see Smelser, 2005). Research on the conditions for scientific advances thus encompasses a large and diverse range of inquiry into the organization and performance of scientific endeavors.

The questions raised by our charge arise across the sciences. For example, many of the questions posed by BSR also arise in the conduct of industrially funded research, have been posed by industrial research and development managers (Industrial Research Institute, 1999), and are the subject of an extensive literature on research and development portfolio selection (e.g., Bretschneider, 1993). These questions are also commonplace in the science priority-setting processes of other countries, as evidenced prominently in the Foresight and related activities (described below) that enter into the formulation of the European Union's Framework programs (Pichler, 2006). To review all the literature on these questions across the sciences and internationally would widen the scope of our inquiry to unmanageable dimensions.

Restricting our search only to literature dealing explicitly with behavioral and social science research on aging, however, would unacceptably narrow the scope of this study. If we used the narrower approach, we would fail to take advantage of research in other areas of science that has lessons

to offer. Thus, we have used our judgment in selecting sources of knowledge that seem to provide useful insight for the tasks facing BSR, providing passing coverage of some and omitting others we deemed less germane. We have concentrated on some widely applicable techniques for assessing the past performance and progress of scientific fields and the prospects for scientific progress, and on frequently discussed variants or alternatives to these techniques. We have also conducted a pilot study using some new and promising bibliometric techniques, customized to correspond to substantive areas of research supported by BSR.

This work leads to the conclusion that all available techniques for assessing the progress and prospects of scientific fields embody significant uncertainties and will continue to do so for the foreseeable future. By itself, this is neither a novel nor a surprising conclusion. It reaffirms similar conclusions offered by both older and more recent undertakings (e.g., National Research Council, 2005c). Similarly, our review of studies of the factors likely to contribute to scientific discoveries reveals broad consensus about major influences—“adequate funding,” for example—but uncertainty and indeed disagreement about their programmatic implications. For example, many sources point to the importance of interdisciplinary teams working in an open “collegial” environment. However, more systematic research is needed before it can be concluded that these lessons apply to the conduct of behavioral and social science research on aging.

Given the inconclusive and open-ended nature of current knowledge, the key practical problem for BSR is how to make wise choices when even the best techniques of analysis give uncertain information. BSR, and most likely other parts of NIH and other federal science agencies, need to establish processes for considering, interpreting, and using assessments offered by different parties—congressional and executive branch decision makers, researchers, user communities, program managers—of the potential contributions of different fields of science toward mission objectives. Good processes can integrate improved analytic techniques as they appear, while ensuring that imperfect measures do not trump good judgment.

Working from a focus on process, we propose a strategy that BSR can use for assessing and comparing the value of research across areas of inquiry. The strategy uses quantitative measures, indicators, and the like to inform judgment rather than to replace it.⁴ It treats analytic techniques, including the application of indicators, as useful for disciplining the judgments of expert groups and focusing their deliberations, but it emphasizes the essential contribution of expert deliberation for interpreting quantitative information and informing strategic decisions about research policy. While recognizing the limitations of existing and emerging methodologies, it sees value in experimenting with promising techniques.

We propose this strategy from a recognition that an assessment strategy

should address both the internal needs for decision making—the agency’s specific mission and the methods it finds acceptable for formulating priorities and assessing progress—and the decision context. Context refers to the structures and procedures of the NIH, NIA, and BSR within which decisions are made and to the distribution of decision-making authority and influence among various actors, including program managers, other decision-making entities in the agency, and extramural scientific bodies and policy actors.⁵

Among the key contextual factors for BSR is the paradigmatic use of peer review and expert judgment mechanisms in NIH, as in other federal science agencies, such as the National Science Foundation (NSF) and the Department of Energy. The context also includes the accepted structures for priority setting and proposal selection in NIH (see Chapter 2). Over time, what is learned from this study and others with similar objectives (e.g., National Research Council, 2005c) may lead to broader understanding of the generic issues that may be useful for assessing science and setting research priorities in other domains and organizational contexts as well.

The tasks for this study involve both prospective and retrospective assessment. Priority setting has a prospective focus. Working in a decision-making context shaped by legislative and executive branch mandates, budget allocations, political imperatives, stakeholder interests, and inputs from the affected scientific communities, agency program managers consider how best to distribute their programs’ available resources among many possible lines of science to maximize attainment of the program’s goals, such as the advancement of scientific knowledge and human well-being.

Assessments are usually retrospective. They may be conducted for summative purposes—that is, to determine how well past funds have been spent—and for formative purposes—to generate lessons learned for future decisions. Thus, retrospective assessments can affect the allocation of future funds across research fields, types of funding mechanisms (such as between individual investigator awards and multidisciplinary centers), and types of recipients (individuals or organizations).

The connection of the past with the future is not always linear or predictive. This is especially the case during periods such as the present when there is a widespread consensus in NIH, as indicated by its Roadmap initiatives (<http://nihroadmap.nih.gov/initiatives.asp>), and other federal science agencies, as illustrated by the National Science Board’s 2020 Vision for the National Science Foundation (<http://www.nsf.gov/pubs/2006/nsb05142/nsb05142.pdf>), that transformational changes are occurring in relationships among scientific and technological fields and increased attention is being given to the need to translate research findings into techniques, methods, and policies that enhance human health and well-being.

In this report, we propose a strategy for assessing the progress and prospects of science that embeds analytic techniques in a structured deliberative

process. We think this strategy will make sense both to science managers and to working scientists involved in BSR's domain of responsibility and that it will allow for discrepancies in judgment between different individuals or groups to be deliberated in a more informed way than in the past. Over time, with examination and reflection on how the strategy is used in BSR's advisory processes, it will be possible to continue to improve practice.

GOALS OF THE STUDY

The BSR Program of the NIA is the lead federal agency assigned the mission of supporting behavioral and social science research related to aging. As described on the NIA web site (<http://www.nia.nih.gov/ResearchInformation/ExtramuralPrograms/BehavioralAndSocialResearch/>), BSR focuses its research support on the following topics:

- How people change during the adult lifespan
- Interrelationships between older people and social institutions
- The societal impact of the changing age composition of the population

BSR support has emphasized "(1) the dynamic interplay between individuals' aging; (2) their changing biomedical, social, and physical environments; and (3) multilevel interactions among psychological, physiological, social, and cultural levels." In pursuit of its objectives, "BSR supports research, training, and the development of research resources and methodologies to produce a scientific knowledge base for maximizing active life and health expectancy. This knowledge base is required for informed and effective public policy, professional practice, and everyday life. BSR also encourages the translation of behavioral and social research into practical applications." NIA expends the bulk of its funds on grants and contracts.

BSR is seeking to address the challenges of research assessment and priority setting explicitly and systematically. It seeks to develop valid and defensible procedures for making judgments about the progress and prospects of the scientific activities it supports at the level of lines or areas of research. It seeks to identify the factors that contribute to discovery so as to have a firmer basis for allocating and reallocating funding across types of funding instruments and types of recipients (e.g., grants for research projects versus programs; grants to individuals versus research groups; disciplinary versus interdisciplinary research teams).

It seeks improved procedures for assessing scientific progress and prospects and firmer rationales for allocating incremental research funds across areas on other than a percentage-based formula and, as appropriate, for reallocating research funds from one area to another. By requesting this

study, BSR has offered itself as a test bed for addressing important generic priority-setting questions that arise in many areas of federal government science policy. One of these is how best to assess the performance of investments in science when some of the objectives of those investments are hard to quantify (e.g., improving knowledge, the quality of policy decisions, or human well-being). Another is how to compare the performance of different kinds of investments when the sponsoring agency has multiple goals and different lines of research contribute to different goals.

A third is how to assess the progress and prospects of scientific fields that differ systematically in their basic objectives, methods, and philosophical underpinnings. The social and behavioral sciences exemplify this issue well. Despite much-discussed trends toward consilience across fields of science and convergence and cross-fertilization among the behavioral and social sciences (e.g., behavioral economics), significant differences in philosophical underpinnings and methodologies remain among and even within these disciplines (see, e.g., Furner, 1975; Ross, 2003; Ash, 2003; Stigler, 1999).⁶ These differences underlie the historical division of the behavioral and social sciences into disciplines and subdisciplines, are unlikely to be easily resolved, and serve as the basis for competitive claims on the support provided by research sponsors, such as BSR.⁷

A fourth issue is the effects of priority-setting decisions by major research funding organizations on the competition among disciplines and departments in the contemporary American research university. Assessments of scientific fields at times become enmeshed in disciplinary rivalries. Indeed, our assessment highlights the challenge to BSR of disengaging its problem or mission focus on aging from the claims of different academic disciplines to “own” a particular facet of research on aging. The progress of disciplines, however measured, does not automatically translate into progress in the kinds of areas of inquiry of greatest interest to BSR or similarly mission-oriented science programs. In this report, we use such terms as research “areas” or “fields” flexibly to refer to topics or lines of inquiry that may be as appropriately defined by a problem as by a discipline or subdiscipline.

The questions that BSR is asking, especially about the comparisons among the several areas of behavioral and social science research it supports, have received surprisingly little systematic attention. Research agencies often engage in serious efforts at priority setting, but comparative assessments of lines of research within or across scientific fields are usually approached indirectly or implicitly. For example, the National Research Council has often been asked to advise federal agencies on criteria for making such assessments (e.g., Institute of Medicine, 1998, 2004; Committee on Science, Engineering, and Public Policy, 2004; National Research Council, 2005c) or to identify priority areas for research from among a broad range of possibilities in many disciplines (e.g., Institute of Medicine, 1991; National

Research Council, 2001b). The typical method for providing an answer involves creating an expert group and asking it, often after considering input solicited from the relevant research communities, to deliberate on the question at hand and arrive at a consensus judgment that is advisory to the relevant decision makers. Only occasionally have such groups been self-conscious about developing and applying explicit methods for comparing fields so as to set priorities among them (e.g., National Research Council, 2005a, 2005c).

Scholarly work on the assessment of science and the operation of scientific advisory panels has focused on somewhat different questions. For example, there has been considerable empirical research on the process of review for individual research proposals (e.g., Cole, Rubin, and Cole, 1978; Cole and Cole, 1981; Cole, Cole, and Simon, 1981; Abrams, 1991; Blank, 1991; Wessely, 1996; Lamont and Mallard, 2005), and some studies aimed at comparing larger scale activities of a single type, such as graduate departments in the same field (e.g., National Research Council, 2003) or research enterprises in a single field but in different countries (e.g., Committee on Science, Engineering, and Public Policy, 2000). Scientists and science policy analysts do sometimes make comparisons among research fields, but seldom in ways that would provide validated decision techniques to a research program manager. Members of scientific communities sometimes disagree about federal agency research priorities, as evidenced by disagreements concerning the budgetary priorities that should be accorded to the superconducting supercollider, the relative emphasis in energy research between discovering new fuel sources or improving energy-saving technologies, and the relative priority of manned and unmanned space exploration. However, research communities typically do not try to resolve such disagreements by applying formal assessment methodologies, such as those of benefit-cost or decision analysis. When challenges are posed to the intellectual substance or vitality of lines of research, they typically are directed at newly emerging ones, particularly those whose conceptual or methodological underpinnings deviate markedly from mainstream fields—and they are focused on attributes of the field in question rather than on techniques for comparison.

One interesting recent exception to these observations is empirical research that is beginning to investigate the characteristics of “successful” interdisciplinary research programs in ways that could help build a knowledge base that could inform systematic comparisons of substantively dissimilar activities or organizations (e.g., Hollingsworth, 2003; Mansilla and Gardner, 2004; National Research Council, 2005b; Bruun et al., 2005; Boix-Mansilla et al., 2006; Feller, 2006). Relatedly, as federal science agencies actively promote interdisciplinary research initiatives, as in the NIH Roadmap, they are beginning to experiment with new procedures for making comparative assessments of the quality of proposals from different fields,

including more deliberate attention to establishing review panels composed of experts from different disciplines (Boix-Mansilla et al., 2006). Although not directly intended as a means of assessing the scientific vitality of different fields or their projected contribution to important societal objectives, the deliberations and conclusions of such panels may provide insights into how to make comparative assessments across fields.

BSR is seeking more systematic methods for such assessments, in part because of a judgment that its interdisciplinary advisory panels have not responded to the issue of comparative assessment of research fields with assessments that differentiated among fields according to the likelihood of returns from research investments. When such differentiation is needed, BSR wants valid ways to justify its recommendations about program priorities and proposal selections to senior NIH officials, Congress, and affected stakeholder and research communities.

The primary focus of this report is on questions of comparative assessment at the level of areas or fields of scientific research. It is not concerned with the overall assessment of the BSR research portfolio in the larger context of NIA or other NIH institutes. Neither is it concerned with comparisons among individuals, research projects, nor university programs, even though some of the methods we discuss have been applied at these levels of analysis. Also, the report's focus is primarily on behavioral and social research, though its analysis and conclusions may be applicable to research in other sciences. Finally, the report's focus is on the needs of an agency whose mission includes both the advancement of basic scientific knowledge and its application to a particular social goal: to improve the health and well-being of older people. An agency with such a twofold mission faces a more complex assessment problem than one whose mission is restricted either to pure science or to specific practical applications of science. In keeping with NIH's overall mission and traditions, it needs both to adhere and advance standards of the highest scientific merit and assess the contributions of fields of science, existing and embryonic, for their potential contributions to NIH's overarching missions.

ORGANIZATION OF THE REPORT

This study has been asked to address several interrelated questions. We have centered our endeavors on what we consider to be the core questions, which concern ways to make defensible assessments of the progress and prospects of areas of scientific research for the purpose of setting priorities among public investments of science within the mission and organizational settings in which BSR functions. In addressing these core questions, we have addressed the remaining questions either explicitly or implicitly. In keeping with our emphasis on context, we begin with BSR's activities.

Chapter 2 considers the BSR Program at NIA. It describes the strategic goals for research in that program and the parent institute, shows the kinds of research investments that have been made, and describes the ways in which the portfolio of research investments is currently evaluated.

Chapter 3 examines what is at stake in research assessment. It briefly reviews the history of federal science priority setting and the debates over priority setting and science assessment, focusing particularly on how concerns with accountability have supported pressures for quantification and the consequent debate over the strengths and limitations of quantitative and other methods for science assessment, particularly traditional peer review. Finally, it addresses the important question of the balance of power and influence between scientists and managers that underlies debates over quantification.

Chapter 4 presents an overview of theories of scientific progress, with special attention paid to the generic problem of the comparative assessment of research fields. It considers what is known about the nature and processes of scientific progress and about the links from societal progress to societal benefit, the variety of kinds of progress that science makes, the factors that contribute to scientific discovery, and the implications of each of the above for priority setting among scientific fields.

Chapter 5 examines the major methods available for assessing scientific progress in a general framework that distinguishes methods that emphasize the use of quantitative measures (analytic techniques, such as the use of bibliometric indicators and the application of decision analysis), methods that rely heavily on deliberation in groups of experts (e.g., traditional peer review), and those that explicitly combine analysis and deliberation. It considers each method and the three general strategies in the context of NIA's objectives, the needs for accountability and rational decision making in science policy, and current knowledge about how science progresses.

Chapter 6 presents the committee's findings and recommendations. It describes our recommended strategy for assessing and comparing the progress and prospects of scientific fields and our specific recommendations for implementing that strategy for assessing the fields of behavioral and social science research supported by NIA.

NOTES

1. Benefits are usually considered in terms of two main kinds of values: expanded knowledge and societal gain. These values are made explicit in proposal review criteria. For example, the NSF identifies two review criteria: intellectual merit (e.g., importance to advancing knowledge and understanding, exploration of creative and original concepts) and broader impacts (e.g., promoting teaching, training, and learning; broadening the participation of underrepresented groups; enhancing the infrastructure for research and education; and benefiting society). The NIH lists five criteria for evaluating applications: significance (e.g.,

importance of the problem, likely effects on scientific knowledge or clinical practice), approach (e.g., adequacy of conceptual framework, research design), innovation (e.g., originality, challenging existing paradigms, testing innovative hypotheses), investigators (their training and suitability), and environment (suitability of the scientific environment for success) (see <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-002.html>). Among these five, the benefits are listed under significance and innovation. In science policy, benefits are also judged against costs, that is, against alternative uses of the funds, and the efficiency and effectiveness with which funds are used.

2. We use “science manager” as a generic term to cover a variety of positions and titles found across federal agencies. Generally, these positions include responsibility for developing intra-agency program and budget plans; maintaining contact with relevant scientific communities; overseeing proposal review and selection processes; endorsing, modifying, or rejecting recommendations made by proposal review panels and justifying these choices to higher organizational levels; and identifying research initiatives.
3. The latter request is a perennial of U.S. science policy. Four decades ago, one of the two questions posed by the U.S. Congress to the National Science and Technology Council (National Academy of Sciences, 1965:1) read as follows: “What judgment can be reached on the balance of support now being given by the Federal Government to various fields of scientific endeavor, and on adjustments that should be considered, either within existing levels of overall support or under conditions of increased or decreased overall support?”
4. We accept van Raan’s (2004:22) definition of an indicator as “the result of a specific mathematical operation with data” designed to serve the purposes of (a) describing “the recent past in such a way that . . . can guide us, can inform us about the near future” and (b) contribute to testing “aspects of theories and models of scientific development and its interaction with society.”
5. The critical role of context in science policy decision making was expressed concisely by Harvey Brooks (1965:99), as follows: “criteria are considerably less important than who applies them . . . [T]he fundamental problem of resource allocation within basic research is who makes the important decisions and how they are made.”
6. Deep philosophical differences exist even within single social science disciplines. Lamont (2004:8) has observed with reference to sociology that it “produces different types of knowledge . . . and that this diversity should be acknowledged in our definition of theoretical growth or vitality. To order sociological contributions within a single hierarchy or paradigm, as economists do . . . would be to weaken it by underestimating the contributions of its various strands. . . . It also would place our discipline very low on the totem pole of fields, which to my view would grossly misrepresent the many contributions of our paradigmatic discipline.”
7. For all the interest expressed by behavioral and social scientists in having a secure and stable home in NIH for basic behavioral science research and training, these communities have expressed little interest in changing the structure or functioning of existing basic behavioral and social science research programs across institutes (Association for Psychological Science, 2005).

2

The NIA Behavioral and Social Research Program

This chapter describes behavioral and social science research in the National Institute on Aging (NIA), the research investments made by the Behavioral and Social Research (BSR) Program in NIA, and the ways in which research investments are currently evaluated, prospectively and retrospectively, in the National Institutes of Health (NIH) generally and BSR specifically.

STRATEGIC GOALS OF NIA

NIA is one of 24 grant-making entities among the 27 institutes and centers that make up the National Institutes of Health. In each institute or center, scientific programs are organized into programmatic areas, such as the BSR Program in NIA. NIA's mission, as described in its strategic plan (National Institute on Aging, 2001; <http://www.nia.nih.gov/AboutNIA/StrategicPlan/>), is "to improve the health and well-being of older Americans through research." NIA supports "research on aging processes, age-related diseases, and special problems and needs of the aged," the training of researchers for work in these areas, resources for accelerating research progress, and dissemination of information on research advances and directions to the public and interested groups. Pursuant to its mission, NIA supports research in a variety of biomedical and social science areas. Box 2-1 lists the institute's major research goals, which are described in greater detail in the strategic plan. Even this abbreviated listing makes clear the breadth of the institute's mandate. NIA implements its mission by supporting both biological and social science research; both problem-focused, applied research

BOX 2-1
**Research Goals and Subgoals of
the National Institute on Aging**

Goal A: Improve health and quality of life of older people

1. Prevent or reduce age-related diseases, disorders, and disability
2. Maintain health and function
3. Enhance older adults' societal roles and interpersonal support and reduce social isolation

Goal B: Understand healthy aging processes

1. Unlock the secrets of aging, health, and longevity
2. Maintain and enhance brain function, cognition, and other behaviors

Goal C: Reduce health disparities among older persons and populations

1. Increase active life expectancy and improve health status for older minority individuals
2. Understand health differences associated with race, ethnicity, gender, environment, socioeconomic status, geography, and culture
3. Monitor health, economic status, and life quality of elders and inform policy

Goal D: Enhance resources to support high-quality research

1. Train and attract a diverse workforce of new, mid-career, and senior researchers necessary for research on aging
2. Develop and sustain a diverse NIA workforce and a professional environment that supports and encourages excellence
3. Disseminate accurate and compelling information to the public, scientific community, and health care professionals
4. Develop and distribute research resources

SOURCE: National Institute on Aging (2001).

and research on basic processes related to health, illness, and well-being; and research at all levels of analysis from the molecular to the societal.

Reflecting the range of scientific research areas supported by NIH, NIA organizes its research support within four programs: Biology of Aging, Geriatrics and Clinical Gerontology, BSR, and Neuroscience and Neuropsychology of Aging. Because of the variety of types of research that NIA supports, its decisions about research portfolios can be quite challenging. Among these challenges, especially in apportioning funds, are those of comparing and setting priorities among research fields. Moreover, the institute's mission requires it to judge research fields both on scientific grounds and in

terms of their potential to improve the health and well-being of older Americans. Thus, its priority setting and budget planning unavoidably involve comparisons of research investments both among program areas and within programs. The questions BSR has posed for this study are integral not only to its mission, but also to those of NIA and NIH.

THE DIVERSE BSR RESEARCH PORTFOLIO

As stated on its web site (<http://www.nia.nih.gov/ResearchInformation/ExtramuralPrograms/BehavioralAndSocialResearch/>), the BSR Program “supports basic social and behavioral research and research training on the processes of aging at both the individual and societal level.” It focuses on the following:

- How people change during the adult lifespan
- Interrelationships between older people and social institutions
- The societal impact of the changing age composition of the population

Emphasis is placed on: (1) the dynamic interplay between individuals' aging; (2) their changing biomedical, social, and physical environments; and (3) multilevel interactions among psychological, physiological, social, and cultural levels.

BSR supports research, training, and the development of research resources and methodologies to produce a scientific knowledge base for maximizing active life and health expectancy. This knowledge base is required for informed and effective public policy, professional practice, and everyday life. BSR also encourages the translation of behavioral and social research into practical applications.

In the 2005 fiscal year, the BSR Program awarded \$159.5 million for research—about one-fifth of all awards in NIA and a doubling from the level of 1997 in current dollars. This included approximately \$75 million in support for investigator-initiated projects (funding categories R01 and R03), \$28 million in support for larger program projects (P01), \$14 million in cooperative agreements (U01), and smaller amounts for other categories of research support (see Table 2-1). Tables 2-2A and 2-2B show the amounts awarded for research in BSR's nine topical areas between 1997-2000 and 2001-2005 respectively, highlighting the breadth of the BSR research portfolio (two tables are needed because the organizational structure of BSR changed after FY 2000). It also shows the shifts in levels of support for research overall and in the different areas.

Overall research support doubled between 1997 and 2003, but it has been essentially unchanged since then. As Tables 2-2A and 2-2B show, research support for demographic research in BSR increased every year from

TABLE 2-1 BSR Grants Awarded in Millions of Dollars (numbers of awards in parentheses) by Mechanism in FY 1997-2005

Category	1997	1999	2001	2003	2005
Research program	\$54 (228)	\$69 (243)	\$ 97 (308)	\$122 (311)	\$125 (315)
Center	10 (28)	10 (22)	10 (22)	12 (44)	13 (39)
Small business innovation research	8 (35)	7 (29)	8 (32)	9 (39)	8 (33)
Career	2 (19)	3 (28)	5 (42)	6 (51)	6 (50)
Training	4 (25)	5 (28)	6 (31)	7 (33)	7 (35)
Interagency agreements	2 (11)	5 (31)	4 (31)	4 (26)	4 (20)
TOTAL	\$79	\$98	\$130	\$161	\$164

NOTE: Totals do not add up to exact amount due to rounding off of figures.

SOURCE: Behavioral and Social Research Program, National Institute on Aging.

TABLE 2-2A Distribution of BSR Program Funds (in millions of current dollars) by Research Topic, FY 1997-2000

Topic	1997	1998	1999	2000
Cognitive functioning	13.4	14.9	16.6	16.8
Personality and social psychology	5.2	4.9	5.2	8.4
Old people in society	5.3	7.5	7.6	7.7
Psychosocial geriatrics	11.5	14.8	17.1	19.5
Health care organizations	11.6	9.5	8.8	8.4
Demography	12.3	14.4	15.2	20.4
Population epidemiology	5.9	6.1	7.6	9.7
Health and retirement economics	13.1	13.8	9.1	11.3
Databases (e.g., Health and Retirement Survey)	0.6	0.6	7.7	9.3
TOTAL	79.0	86.6	94.9	111.3

TABLE 2-2B Distribution of BSR Program Funds (in millions of current dollars) by Research Topic, FY 2001-2005

Topic	2001	2002	2003	2004	2005
Behavioral medicine	31.8	44.9	44.0	44.9	42.1
Cognitive aging	20.4	21.7	20.7	22.5	19.6
Psychological development	12.9	15.8	14.6	14.9	20.9
Demography	22.1	30.6	36.3	34.7	28.2
Epidemiology	10.9	7.2	13.0	14.3	15.1
Health and retirement economics	9.5	10.2	11.4	14.7	17.1
Health and social institutions	11.8	2.6	3.6	4.1	3.4
Behavior genetics	0.3	0.6	1.3	1.8	1.8
Databases (e.g., Health and Retirement Survey)	9.6	12.1	12.9	10.3	11.3
TOTAL	129.4	145.7	157.8	162.3	159.5

NOTE: The organizational structure of BSR changed after FY 2000, so that comparability across the full-time period is only modest in some areas. Also, some projects, particularly the larger program projects, have become more interdisciplinary over time and harder to fit into these categories. NOTE: Totals do not add up to exact amount due to rounding off of figures.

SOURCE: Behavioral and Social Research Program, National Institute on Aging.

1997 to 2003, from about \$12 million in 1997 to about \$36 million in 2003, but has since decreased, to about \$28 million in 2005. Support for research in health and retirement economics fluctuated between \$9 and \$14 million per year between 1997 and 2003, and then increased to about \$17 million in 2005. Support in other fields has had other historical records.

Not shown in the table is the trend of funded research, particularly the larger program projects, toward becoming more interdisciplinary over time. This makes the classification of research support into fields somewhat problematic and implies that the time trends for particular research areas should be interpreted as approximations.

The portfolio of research supported by BSR is diverse, in terms of both substantive focus and the ways the research is expected to generate scientific progress and improve the health and well-being of older people in America. This diversity reflects the inherent complexity of understanding behavioral, social, and economic processes associated with aging; the multiple ways in which the dynamics of the aging population of the United States affects individuals, families, communities, and public policies; and continuing debates within and across scientific fields about the causes and impacts of processes of aging and about public policy choices that BSR-sponsored science can inform. As noted in a recent compendium of studies on aging, health, and public policy, “We cannot plan for population change or design appropriate and effective responses without understanding, for example, the processes that underlie increases in longevity, the mechanisms that accelerate or delay the onset of disability, the incentives that affect retirement decisions, including employment and saving for retirement, and the role of public programs and policies in all of these factors” (Waite, 2004:4).

A few examples taken from recent BSR reports of “scientific advances” illustrate the diversity of the topics on which BSR supports research as well as the variety of ways in which BSR support advances the mission of NIA “to improve the health and well-being of older Americans through research.”

- *Understanding causes of longevity:* Studies showing that late child-bearing has positive effects on survival of the oldest old shed more light on the likely longevity of populations now in middle age (Zeng and Vaupel, 2003). Other studies are continuing to explore genetic factors responsible for longevity, using nonhuman models (e.g., Spencer and Promislow, 2005).
- *Health care expenditures and health outcomes:* Research documented the relationship of health status at age 70 to future health care expenditures and raised issues about the future of Medicare (Lubitz et al., 2003; Cutler, 2003). Research from a repeated survey was used to forecast the future nurs-

ing home population (Lakdawalla et al., 2003). Other research documented the relationships of health care expenditures to health outcomes for older Americans and compared them in different regions (Fisher et al., 2003a, 2003b).

- *The course and consequences of cognitive changes in aging:* Studies have demonstrated that social activities, such as playing games and participating in volunteer work, slowed the rate of cognitive decline in people over age 70 (Glei et al., 2005). One study has shown that normal aging-related cognitive declines leave people susceptible to believing claims, such as those in advertisements, even if they are told the claims are false (Skurnick et al., 2005).

- *Health disparities among social groups:* Studies have explored possible explanations for the association between socioeconomic status and mortality and considered the policy implications (e.g., Wong et al., 2002; Adams et al., 2003; Adda et al., 2003; Deaton, 2002). Studies have quantified health disparities among the aging by gender, race, and ethnicity (e.g., McKenna et al., 2005), demonstrated effects of living in poor neighborhoods (Wen et al., 2005), shown that a higher educational level of children is associated with better outcomes of illnesses in their aging parents (Zimmer et al., 2002), and shown differences in primary care physicians' treatment of aging women and men reporting symptoms of cardiovascular disease (Arber et al., 2006).

- *Health effects of emotions:* Studies have demonstrated and begun to explain the adverse health effects of negative affect and negative perceptions of aging (Rosenkranz et al., 2003; Levy, 2003). Other studies have begun to demonstrate explanatory physiological mechanisms, such as a strong association between loneliness and higher blood pressure (Hawkey et al., 2006; Boomsma, 2005) and an association of involvement by men in social networks and plasma fibrinogen concentrations (Loucks et al., 2005).

- *New theories of the aging process:* Research has developed and tested a new evolutionary theory of aging that offers an explanation of the factors that have contributed to low fertility rates and to the consequent aging of human populations (Lee, 2003).

- *Declining rates of disability among older Americans:* Studies demonstrating a declining rate of disability among older populations were analyzed for their implications for the future of Medicare (Freedman et al., 2002). Further studies have suggested that the increasing prevalence of obesity may reverse the decline in disability at about the time when a rapid increase is expected in the Medicare-eligible population (Chernew et al., 2005; Goldman et al., 2005; Lakdawalla et al., 2005; Olshansky et al., 2005; Preston, 2005; Reynolds et al., 2005).

- *Personality changes in aging:* Research has shown that age cohort

is a major factor accounting for change during aging for some personality traits, such as neuroticism, which can affect health outcomes (Mroczek and Spiro, 2003).

- *Effects of early-life stress on longevity:* Research on a human cohort showed that health in early life, including exposure to serious infectious diseases, affects mortality in later life (Costa, 2003). Experiments with worms demonstrated that the expression of a specific gene mediated a positive effect of early-life stress on longevity (Rea et al., 2005).

BSR-funded research focuses on many kinds of outcomes that are important to older Americans: longevity, health and disability, economic security, caregiving, cognitive functioning, and access to health care, among others. Findings from these studies are designed to help older Americans in various ways, from changing national policy to suggesting ways that everyday behaviors can improve well-being in later life.

Methodologically, BSR supports research that varies from applied analyses of secondary data (e.g., modeling future nursing home populations from past data) to theoretical development of an evolutionary theory of aging. The applied research is often directed at near-term, targeted policy and program questions; the theoretical research contributes first to fuller understanding of the dynamics of aging and then to new or reformulated approaches to societal responses to aging.

For this study, an important aspect of BSR-supported research is the great diversity of its disciplinary content. As is evident from Table 2-2, BSR supports research in a number of disciplines and fields, including psychology, demography, economics, sociology, and population epidemiology, among others. Increasingly, its support has gone to interdisciplinary research. One of many such examples is research drawing on concepts from sociology, social psychology, and physiology that demonstrated how the stress of low-status jobs could trigger metabolic mechanisms that predispose people to cardiovascular disease (Brunner et al., 2002). In this respect, BSR research support mirrors current trends in NIA and NIH, which increasingly reflect appreciation of the interdisciplinary structure of scientific inquiry and the contribution that an interdisciplinary orientation can make to the relevance and application of research findings to practice and policy (National Research Council, 2000a, b).

BSR-sponsored activities have also led to the development of new interdisciplinary areas of research by catalyzing direct interaction among researchers working on similar issues from different disciplinary perspectives. For example, an April 1996 workshop that brought together demographers, evolutionary theorists, biologists, anthropologists, and others to share their understandings of human longevity led to a report (National Research Council, 1997) that named and help create the new interdisciplinary field of

biodemography of aging (see Box 2-2 for more on this field). Developments in the new field led to calls for collecting social and biological data in the same survey instruments. A workshop in February 2000 brought together a broad interdisciplinary group, this time including economists, ethicists, pathologists, and others, to consider the potential and the risks of gathering biological data along with social survey data and to discuss what kinds of biological data would be most useful to social scientists studying processes of aging. The report of this workshop (National Research Council, 2001d) helped further advance interdisciplinary research linking the biological and social influences on processes of aging.

In addition to supporting research that produces scientific results such as the above, BSR invests in the development of databases and measurement techniques that make these and other studies possible and in workshops and conferences that bring together researchers from different fields. These investments are designed to stimulate research and discussions in order to sharpen research questions in existing disciplinary fields, consider the applicability of new analytical methods and databases, explore and refine research questions at the interstices and overlaps among disciplines, and open up new fields for research. BSR supports the development and archiving of multidisciplinary databases, such as the Health and Retirement Study, the National Long Term Care Survey, the Wisconsin Longitudinal Study, and many others (for a list, see Behavioral and Social Research Program, 2004), that make it possible to conduct research that crosses disciplines. These databases allow exploration of important issues germane to NIA's mission, such as the effects of socioeconomic status on health and the causes of declining disability among older Americans.

BSR-funded research also provides useful knowledge for informing public policy decisions. For example, continuing research on declining physical disability and cognitive impairment among older Americans is relevant to policy decisions about the future of Social Security and of Medicaid, which pays for almost half of all nursing home costs. Estimates of the magnitude of decline are directly useful for anticipating the future costs to governments of caring for disabled older adults and thus to financial planning for Medicaid. Also, to the extent that disability decline is caused by events in people's life experiences, research that identifies the causal factors may suggest small investments in well-being that will have large future payoffs in declining financial costs for health care and improved well-being of older adults and caregivers.

BSR values research that bears on any of a broad range of aspects of health and well-being. Thus, the program applies disjunctive outcome criteria in making retrospective assessments of the value of the research it supports. It values both intellectual and practical results, but it does not expect that any particular study or area of research must produce both to

BOX 2-2 **Biodemography of Aging**

BSR currently supports work in the field of biodemography of aging through two program projects: one at the University of California, Davis, on the “Biodemography of Life Span” (e.g., Carey, 2003; Carey and Tuljapurkar, 2003), and one at Duke University on “Oldest Old Mortality” (Vaupel et al., 1998). It also supports a number of smaller research projects in this area. Biodemography, the confluence of biological and demographic studies of aging, has made important contributions in several areas. A significant contribution of BSR-supported work in this area is that it provides the central link between demographic and epidemiological studies on humans and detailed biological studies of other organisms.

BSR has supported comparative empirical work on several species (fruit flies, Medflies, yeast, nematodes, and humans) that has demonstrated characteristic patterns of old-age mortality across species in which the mortality hazard becomes roughly constant at high ages. This is an important finding and influences much current work on mortality at the oldest ages (100+) in humans. This finding has also focused attention on the biological and evolutionary determinants of mortality trajectories (Vaupel et al., 1998).

BSR support has stimulated new and important work on the evolutionary basis of mortality patterns. Classical theories of senescence view mortality patterns as a result of a balance between deleterious mutations and selection that eliminates such mutation. Work by Lee (2003) shows that evolution of mortality in social species can be rather different—sociality in many species results in intergenerational transfers of time, food, and care (e.g., from women to their grandchildren), and these transfers can alter the strength of selection on old-age mortality by making older individuals more important contributors to fitness. Other recent work has examined the relationship between mutations and selection in shaping the age trajectory of mortality and has greatly

be valuable. It also values efforts such as databases, workshops, and conferences, which provide lines of communication or infrastructure that can result indirectly in improved theory or applied results. In making prospective judgments about research, BSR is similarly eclectic and open to the many kinds of benefits that research can provide.

DECISION-MAKING PROCESSES AT BSR

With so many objectives and so many ways that social and behavioral research contributes to the understanding of processes of aging and the

extended population genetic theory in this area (Steinsaltz et al., 2005). All these lines of research are shaping the search to understand the specific genetic and environmental determinants of mortality in humans.

BSR has stimulated research on the heritability of longevity using twin register data. Researchers in the Duke project (Manton and Yashin, 2000; Christensen et al., 2001a, 2001b; Tan et al., 2004) have developed valuable methods for integrating genetic and demographic data to distinguish genetic from environmental influences on mortality. They have found that variance in longevity has a significant, but not a predominant, genetic component—not much larger than is typical for other quantified traits. This finding suggests that the key to understanding longevity does not lie in a few genes and implies that the analysis of mutation-selection dynamics in determining longevity is likely to have a high payoff. Curtsinger and colleagues (Pletcher and Curtsinger, 2000a, 2000b; Khazaeli and Curtsinger, 2001) have produced similar findings in *Drosophila* species and are now making progress at identifying the quantitative trait loci that are responsible for shaping differences in longevity within and between populations.

BSR has supported detailed studies of how proximate factors, such as physiology of metabolic and reproductive energy use, dietary variability, and timing of reproductive and sexual activity, affect mortality in several species (Cargill et al., 2003; Carey, 2003; Carey et al., 2005; Vaupel et al., 2003). These studies have produced useful methods of analysis that are applicable to human studies, and the substantive results show how physiology and diet interact with reproductive metabolism to shape mortality. At this stage, these studies have raised as many questions as they answer, but they are key in designing future experiments and in providing mechanistic ways of studying such phenomena as the Barker effect—namely, that early life (fetal and infant) conditions can have significant effects on mortality in later life (Barker et al., 2002).

well-being of older Americans, BSR is routinely faced with the tasks of allocating money efficiently within and across different research fields and among research applications. Changes in the national conditions of older Americans, government priorities, and science also require that BSR continually reassess these allocations. The tasks become more challenging when research funds tighten in relation to the number of proposals being received and the number of researchers working in the fields in which BSR supports research. This has been the case in BSR and NIA more generally since 2003, as it has been across NIH (Mandel and Vessel, 2006). The success rate of applications in BSR, that is, the percentage of applications reviewed that are

actually funded, has declined from 30.7 percent in FY 2001 to 17.1 percent in FY 2005. There have been similar declines in other units in NIA (National Institute on Aging, 2005).

Concern with Relative Quality

Ferment exists in several federal science agencies about the adequacy of existing priority setting and proposal selection procedures for selecting high-payoff, transformative, and high-quality research proposals. Expressed with varying degrees of explicitness, concerns about undue conservatism, resistance to interdisciplinarity, and the existence of unacceptably low field-specific or study group-specific standards of quality inherently reflect criticisms of the discipline-based peer review system. This report responds to the specific concerns of the director of BSR regarding the adequacy of the program's current decision-making processes for assessing research proposals across fields and for assessing its overall research portfolio.

One concern affects the assessment of research proposals, specifically the possibility that different standards of research quality are being applied to proposals in different fields, some of them "higher" than others. Imagine a scale of height in which some people measure in inches and others in centimeters, without anyone knowing what units others are using, and in which items are compared by the numbers without regard to the units. If that is an apt analogy for the rating processes used at BSR (or elsewhere in NIH), in a community-dominated decision process, proposals of high quality in high-standard (inch) fields would be denied support while lower quality proposals in low-standard (centimeter) fields would be funded. To put this idea in the language of ranking, the concern is that proposals judged to be in the top 10 percent in quality in one field might have much lower ranking according to the standards of another field.¹

A related concern operates at the level of portfolio decisions. To the extent that scientific advisory groups contain people who apply different standards to research, group consensus processes may be counterproductive, perhaps by leading to support for the least objectionable proposals or to schisms within advisory groups that can have unpredictable effects on their advice. If advisory group members with different standards trust each other in their narrower areas of expertise, the group's advice would be to support more research in the weakest areas than would be warranted on an objective quality standard. If members do not trust each other, they would have no way to reach consensus advice on the recommendations to offer to program managers (Brenneis, 1994). One result might be that a group offers consensus advice based on tacit agreement to trust each expert's judgment in his or her own area, while individual advisory group members grumble in private that the group has endorsed supporting weak research.

Some researchers, including leading researchers, have expressed discontent in this regard about existing study sections and review panels in both NIH and the National Science Foundation (NSF). The challenge for science assessment is to determine whether such expressions derive from the kind of problem described here, or instead reflect the idiosyncratic, parochial, or selective disciplinary interest.

As expressed here, these concerns presume that there is a single quality metric that can be applied to research across scientific fields: that differences in standards are differences in the ways a single underlying dimension of quality is being assessed, not differences in the kinds of things being assessed or in judgments of the relative importance of different objectives (such as scientific quality and potential societal benefit) that are supposed to be combined in making assessments.

The above discussion of BSR-funded research suggests that the program does not in fact apply a single metric in assessing the results of the research it supports. At least in terms of the ways that research contributes to the “health and well-being of older Americans,” several modes of contribution seem almost equally important. In this respect, BSR appears to apply disjunctive criteria of value, such that a research activity may be judged as having resulted in an advance or a discovery if it satisfies any one or two of a relatively long list of implicit criteria.

Our collective experience with BSR, coupled with the expressed concerns of program managers and leading researchers, lead us to take seriously the possibility that existing priority setting and proposal selection processes may not be doing an adequate job of ensuring that the program supports research of a uniformly high “quality” across the fields the program supports. We take these concerns seriously even though it seems clear that quality is not being judged and should not be judged on a single dimension. Addressing concerns about relative quality requires developing a more nuanced idea of the dimensions of quality that might appropriately be used to evaluate research portfolios. Here we suggest a few dimensions of research quality that might apply in BSR and consider which of them might provide cause for concern. In Chapters 3 and 4, we address knowledge about the progress of science and ways of assessing the outputs and potential value of science when criteria are multidimensional.

Dimensions of Quality

For a research sponsor such as NIA, various dimensions of research quality might be used for prospective or retrospective assessment. We briefly note a few.

- *Conceptual and methodological quality:* Scientific research is often

judged by how clearly its concepts are defined, how reliably they can be measured, and how well the operational measures of variables and concepts correspond to the underlying concepts themselves. Scientific progress sometimes consists of developing clearer concepts when they had been ambiguous, better measures when they had been unreliable, or better evidence that a measure is a valid indicator of a construct of interest. Thus, one appropriate way of judging the quality of research is methodologically in the terms of the research itself: research passes a quality test if it uses clear concepts and reliable and valid measures. It passes another quality test if it moves scientific discussions in its field toward using clearer concepts or better measures.

- *Advancing thinking:* Scientific progress is often defined in terms of efforts to develop and validate theories or conceptual models that attempt to make sense of whole classes of phenomena and thus provide explanations of some generality. If a research activity develops a new theory to explain phenomena that were previously thought to be unrelated, or develops a new way to test a theory, or produces results that call an existing theory into question, its quality may appropriately be judged higher than a research activity that simply replicates a previous finding or is presented in a way that is unrelated by theory to any wider set of phenomena. Of course, the fact that a piece of research calls previous findings into question does not imply that it meets other quality criteria.

- *Generating widely applicable research findings:* Scientific research is also judged by the extent to which its results stimulate researchers, particularly outside its immediate field, to carry out new research. For example, the development of game theory as an analytic method has led to a variety of applications to problems in economics, social psychology, international relations, and other fields. BSR-sponsored research linking social and emotional factors to illness and longevity (e.g., Brunner et al., 2002; Adams et al., 2003; Rosenkranz et al., 2003) may stimulate various lines of research to elucidate the mechanisms explaining these relationships, with possible applications to more effective health promotion.

Research can generate widespread interest in various ways, including developing new lines of theory, raising new research questions, calling into question widely accepted societal beliefs, and developing new ways to study understudied phenomena. Even research that seems at first to be of interest to only a few academics may later become the foundation stones of far-reaching theoretical and applied advances. Generally, though, the more widely cited research findings are, not only within but across disparate fields of investigation, the more important those findings tend to be.

- *Practical application:* Science that is supported for its potential benefits for health and well-being, as at NIH, can also be judged on the basis of its actual or potential practical application—its “broader impacts,” in

the language of the NSF merit review criteria (available: <http://www.nsf.gov/pubs/1999/nsf99172/nsf99172.htm>). Research can be appropriately judged to be more important on this dimension from an NIH perspective if it produces (or can reasonably be expected to produce) results, methods, techniques, or other outputs that advance the sponsor's ultimate societal mission. Research that is unlikely to produce such results can be considered to be less important on this dimension. As already noted, a wide variety of possible practical applications are of interest to NIA. Research may be rated highly in terms of practical application if it can be applied to any important aspect of health or well-being and if the applications are likely to affect the health or well-being of large numbers of people.

Considering the variety of dimensions on which research quality may be judged, it is reasonable to ask which dimensions are of greatest concern to BSR in making comparisons of research in different fields. This is, of course, a policy question for BSR and NIA. Comparative assessments of fields will be more responsive to the sponsor's needs if BSR can be specific about the objectives it deems most important to consider in assessing the relative contributions of research in different fields. With clarification of the objectives, it becomes more possible for BSR to identify possible metrics for the objectives or, when there are no generally agreed-on measures, to move expert deliberations toward making more explicit comparisons along those dimensions.

Comparisons across research fields should consider the quality and importance of research that has been supported in relation to the levels of support that have been made available for research in the fields. As can be seen from Table 2-2, a considerable range exists in funding levels among BSR's branches, from less than \$2 million in 2005 for research in behavior genetics to about \$42 million for research in behavioral medicine.

RESEARCH REVIEW IN NIH

NIH research review practices set a context for any effort to reconsider the practices for setting research priorities. They do this in at least two ways: they affect the pace at which review and advisory panels and their decision rules can change, and they provide opportunities and constraints for the discretion of research managers.

Funding and Peer Review

Extramural funding at NIH is made primarily through grants and contracts. Contracts are used when the agency defines the work to be done and intends to use the product of the contract. Grants are based on the

premise that there is a shared interest in the outcomes of the research. A third mode, cooperative agreements, involves grants in which there is an understanding that the federal sponsor will be involved in a predetermined way in the work of the grant. Review of research proposals at NIH follows different procedures depending on the funding instrument and on whether the proposals are solicited or unsolicited.

The largest component of most programs, including BSR, consists of projects funded through grants as a result of unsolicited, investigator-initiated submissions. Researchers submit applications to NIH, and they are assigned by the Center for Scientific Review (CSR) to a relevant institute or to multiple institutes for administration. Within the institute(s) they are assigned to programmatic areas, such as BSR. These assignments are based on referral guidelines, updated periodically, in which institutes and programs have outlined their scope of interests. These grant applications are also assigned to review groups, almost always established study sections, for merit review. Study sections cover a defined range of substance that may fall within the purview of a single institute or center or cut across many.

Members of review groups are proposed by the scientific review administrator in CSR and approved by NIH. Membership is typically for four years, and therefore the composition changes by about 25 percent each year. In addition to regular members, individuals may serve on an ad hoc basis for a meeting or two. This is done to ensure there is a specific skill needed for the review of some application, or a way to try out an individual before nomination for regular membership. Members whose terms are expiring may be asked to suggest possible members to fill their roles, a practice that could lend considerable stability to study sections. However, review groups do change over time as they respond to the changes in the nature of the applications being referred to NIH: if there is growth in a field or area, more reviewers are added to address that need; if applications decline, reviewers in that area become less prevalent on the committee.

In its selection of members of peer review groups, NIH acknowledges the person's stature in a field and its confidence that the individual can provide useful assessments of scientific merit. The members of a group may be called on to review a variety of projects in different fields, so groups are constituted to bring relevant skills to bear on all applications. Applicants do not know which reviewers were specifically assigned to their applications. Panels make two types of recommendations: first, whether a proposal meets threshold levels of scientific quality and conforms to other eligibility criteria; second, assessments of scientific merit, expressed in terms of a numerical "priority score."² Proposals are placed in a rank order based on their priority scores and are normally funded in that order until all or a predetermined percentage of the program's requested budget is committed.

The recommendations of peer review are presented to the relevant

institute's advisory council. The councils have the right to submit their own recommendations. They can recommend against funding projects with higher priority scores and in favor of funding projects that have received less favorable scores on the basis of judgments about program relevance. The institute in turn acts, informed by those recommendations. If there is a question about error in a particular review, the applicant can ask that the program manager consider the applicant's concerns. If the program manager cannot resolve concerns, they can be presented to the advisory council.

Although program management has discretionary authority to forward recommendations that differ from those of the study sections, NIH has built a reputation for adherence to recommendations of scientific peer review. By reputation, program managers infrequently seek to make decisions that deviate from the assessments of the study section. It is not unusual to find that the first 80 percent of the available funds are allocated according to the score, but that for the remaining 20 percent, program manager judgment is factored into the decision whether or not to fund. The typical practice in NIH institutes is to require a strong rationale for any decision by a program manager to fund a project that did not score well or not to fund a project with a high review score. In the latter case, for example, a conclusion that a project duplicates one already being supported provides a sufficiently strong rationale.

NIH program managers are generally seen to exercise less discretionary authority relative to peer reviewers than their counterparts at NSF, and considerably less than their counterparts at the Department of Defense's Defense Advanced Research Projects Agency (DARPA), where program managers, even though they consult widely with experts in a field, are generally viewed as having wide latitude to select from among the proposals they receive.

NIH periodically reviews the structure of review groups and adjusts their scope and the expertise needed. The panel on scientific boundaries for CSR began in 1998, and updates are posted on the CSR web site (<http://www.csr.nih.gov/review/irgdesc.htm>).

The one-quarter annual turnover of review groups could possibly act as a brake on change. Review panels might not quickly reflect changing interests and priorities in the scientific community, and they might continue to provide a favorable home to areas of science after research productivity has peaked. Also, a review group is likely to have few reviewers in any new field. Thus, as is implicit in NIH's articulated rationale for establishing its new Roadmap and for forming new panels to review proposals submitted under this initiative, investigators in a new field may find they are being reviewed by people who do not know the field well.

Changes in the review process can be achieved by revising criteria for scientific merit. For example, in 1997, the criterion of "innovation" was

added to help respond to the concern that review groups had become too conservative (see Notice 97-010 and NIH OD-05-002). Reviewers are required to address how a proposal is or is not innovative, although it is recognized that some meritorious projects may not be innovative. Specific review criteria may be issued for a Request for Applications or Program Announcement solicitation, and special criteria may also be established for certain types of research, such as that involving the preparation of natural products (see Center for Scientific Review notice, 23 July, 2003) or the use of human subjects, children, animals, etc.

Discretion of Research Managers

As already noted, mainstream practice in NIH is for program managers to adhere closely to the recommendations of study sections or ad hoc peer review groups regarding the funding of unsolicited proposals that have been received. Doing otherwise risks raising objections that such actions constitute an erosion of support for peer review.

Program managers, however, have more discretion outside the realm of unsolicited proposals. Acting as scientific entrepreneurs, program managers can encourage research in new fields by soliciting proposals both from existing cadres of researchers and by seeking to enlist the interest of researchers who traditionally have not sought support from NIH. Building a new field of science and building a new community of researchers thus becomes a joint, mutually reinforcing process. An institute has the greatest discretion in contracting for work that meets very specific needs. It may issue a Request for Proposals (RFP), in which it can specify in advance what it wants to have done and whether the product of the contract is for the institute's use. For example, an institute or program can take the product of a contract, such as a data set, and make it available to others to analyze. Contract proposals are typically reviewed by review groups specially created for the purpose and managed by the institute or center rather than by CSR.

Program managers may also shape their research portfolios through Requests for Applications (RFAs) that define a specific problem, goals for the research, and sometimes specifics of methods to be used, such as reanalysis of available data or collection of new data, while allowing latitude in how researchers address the problem. RFAs may include specific review criteria and may also include a set-aside of funds, making clear the institute's commitment to the area. Most RFAs are reviewed by a special review group that encompasses just the applications submitted in response to the RFA. The review may be managed by CSR or by the institute's review office. Plans for RFAs must be reviewed and approved by an institute's advisory council as part of the institute's program plan. They often result from specific consultation with the scientific community. In some cases, an external review

of a field that identifies research needs may be the basis for an RFA (e.g., a National Research Council report). Nevertheless, program managers have latitude in drafting the RFA and in contracting for reviews of particular areas that might lead to subsequent RFAs.

Institutes and centers may also issue Program Announcement (PAs) that remind the scientific community of continuing interest in an area. PAs must be reviewed after three years and either dropped or reissued. PAs may involve a set-aside of funds and may involve a special review, but such special treatment is not necessary. In many cases the responses to a PA are reviewed by the standing review group to which that application would have been reviewed had there been no PA, and funded according to the overall funding plan. If there is a special review, it can be managed by CSR or the center or institute that made the announcement.

Thus, the PA and the RFA are two powerful and efficient ways in addition to contract research that NIH communicates about priorities and interests with the scientific community. All solicitations are published on the NIH web site, and there is a weekly email listing of all announcements issued that week.

RFAs and PAs identify the funding mechanisms to which they apply. They may be restricted to small grants to individual investigators. They may use a mechanism such as the R21 (exploratory/development grants) to show interest in eliciting innovative projects. R21 awards are usually given to researchers who have little or no preliminary data. They have upper limits on length of award and funding, but they provide sufficient funds to get an investigator started on a new line of inquiry. BSR/NIA has announced that it will accept R21 applications in all seven of its major emphasis areas, identifying 18 specific topics of interest (<http://www.nia.nih.gov/ResearchInformation/ExtramuralPrograms/BehavioralAndSocialResearch/R21Grants.htm>). This form of communication with the scientific community of the office's interest in supporting innovative work on certain topics is less formal than the issuance of an RFA or a PA.

An RFA may solicit proposals for large grants such as program projects (P01) or centers (P30, P50), which provide core support or help institutions bring together elaborate teams of people to address a problem. Large grants serve to focus an institution's activity around a research question; the downside to such awards is the possibility that they create vested interests in continuing the support in a recipient institution, sometimes backed by a member of Congress. Thus, it may be difficult or painful for an institute to terminate support for such projects. (Procedures and criteria for evaluating centers and related large awards are reviewed in Institute of Medicine, 2004).

The above discussion illustrates the various means that programs and program managers in NIH can use to grow specific program areas, to shape

a research portfolio, and to provide support to areas that are seen as needing additional attention. In addition, research managers may use formal announcements as a way to bring relevant work together for direct competition rather than having projects come in at different times. BSR has used the full range of options available for identifying new and promising research areas and for encouraging scientists to develop those lines of research.

NIH does not offer many tools to its research managers to exercise discretion in identifying research areas that are not, or are no longer, programmatic priorities. In a tight budgetary context, an institute may take “negotiated” reductions in grants and, in doing so, may make greater reductions in grants in substantive areas in which there is less interest or perceived payoff. Some institutes and centers give program managers latitude in making these judgments. For example, an institute or center might declare that, unless there are specific programmatic or policy reasons, applications will be paid up to 85-90 percent of the funds available. The remaining funds may be allocated to meet programmatic goals without special action of the advisory council. This strategy is typically used when the grants involved are close to indistinguishable in quality according to reviewers’ judgments. Another discretionary tool is that program managers can elect to decline to have a project assigned to the institute or center if it is over \$500,000 in direct costs and does not fit their program priorities.

Institutes, of course, can discourage proposals in an area by simply reducing the amount of funding they provide to it and by publicizing the low absolute and relative success rates of proposals in the area. An institute may also communicate a desire to spend less in a given area by omission: it can publish areas of interest and leave some areas off the list. Expressions of interest (or disinterest) tend to circulate quickly in the scientific community and may influence decisions about research to propose. However, these processes may produce mixed signals and misunderstandings between program officers and researchers about whether it is fields of study that are being deemphasized or specific proposals. In general, though, it is not easy or popular to declare an area of inquiry to be of low interest, so a research manager needs a solid basis and organizational support for such a declaration.

NOTES

1. This concern may play out differently in different review processes. In NIH study sections, members normally come from several disciplines, even if they have in common concern and expertise in a particular research area or field. If researchers in different disciplines have different standards in these terms, the differences in standards are unlikely to perturb the overall decision process unless panel members habitually defer to the judgments of panel members rating proposals from their own disciplines.

2. Reviewers in NIH study sections rate proposals that are of acceptable quality on a scale from 1.0 (outstanding) to 5.0 (acceptable). The average rating is multiplied by 100 and is called the priority score. Priority scores are ranked within the review group to allow the proposals to be prioritized. Based on its available funding, an institute decides what percentage of proposals can be supported and establishes a “payline,” which is a percentile number such that proposals ranked at that percentile or higher among all those reviewed are normally funded. For continuing review groups, percentiles are normally calculated on the basis of the past three rounds of ratings. Funds are normally granted in percentile order, with funding going to proposals scoring above the payline regardless of which study section conducted the review. One effect of this procedure is that study sections are normalized to account for any systematic differences among them in how members rate proposals.

3

The Stakes in Research Assessment

This chapter endeavors to make explicit what is at stake in the comparative assessment of research fields, retrospectively and prospectively. It takes as given that core decisions concerning the scope of any research agency's mission, overall budget, and dominant mechanisms for selecting research proposals are largely determined at high political and administrative levels, such as those of the Office of the President, Congress, and politically appointed agency heads. Science managers in subagency divisions or programs such as the Behavioral and Social Research (BSR) Program at the National Institute on Aging (NIA) work to set priorities within these bounds by identifying, supporting, maintaining, and nurturing scientifically vital, mission-relevant areas of research, including selectively promoting areas of research in the manager's mission area on the basis of judgments of their prospects for advancing that mission. Science managers may also recommend changes in larger organizational priorities and practices deemed necessary to permit specific programs to more effectively achieve their objectives in the context of the larger agency mission. Managers do not do these things on their own, however. Their actions are informed and influenced by outside constituencies, prominently including working scientists in the relevant fields and the potential beneficiaries of the science.

This chapter addresses two major, interrelated themes in science assessment: the actors (Who should be involved in assessing the science, and what should be the relative power and influence among these actors?) and the methods (How should assessment be done and decision making organized?). These questions and themes are being raised in the National Institutes of Health (NIH) and in many other federal government science agencies.

We begin by briefly describing the historical context of science priority setting in the U.S. federal government, which in recent years includes increasing pressure to move from traditional, peer review–based approaches for setting research priorities and assessing returns from research investments to approaches that rely more on quantifying returns from investment in science. We discuss the limitations both of traditional expert judgment and of quantitative approaches, recognizing the particular difficulties of comparing different kinds of fields and of assessing scientific progress in interdisciplinary or transdisciplinary fields. Finally, we discuss the ways in which debates about the best methods for priority setting in the context of a movement for government accountability raise deeper questions about the balance of influence and power among researchers, program managers, advisory councils, extramural scientists, and other interested parties.

These debates occur in the pursuit of two legitimate and important public policy goals: making public expenditures accountable to the taxpayers and ensuring rational priority setting among research expenditures, based on the best available information about the likely returns from future public investments. Accountable and rational priority setting implies a need for comparative assessment, especially when continuing and new claims on resources outpace the rate of increase in science budgets or when a perception emerges that some research fields are not advancing, despite continued support. Comparative assessment is inherently difficult, however, for several reasons. First, different fields of science may produce different kinds of benefits and may benefit different people. In a diverse society, it is unlikely that everyone will agree on the relative importance of different kinds of benefits and therefore on the overall benefit of any particular line of research (Bozeman and Sarewitz, 2005).

Second, some kinds of benefit are easier to see and measure than others. Bibliometric data, for example, make scientific publications and published citations to them highly visible, but they have serious limitations, as detailed in Chapter 5. These data primarily reflect communication patterns among scientists. They are not necessarily predictive of the absolute or relative value of scientific outputs as sources of information for or influence on policy makers.

Third, even if agreement can be reached on the relative benefits from different fields of science to date, the benefits of future investments would remain speculative and uncertain. The existence of compendia of erroneous predictions by experts about the progress of fields of science and technology is good reason to start from the premise that no science of predicting the future of science exists (Cerf and Navasky, 1984, Chapter 7; Thomas, 1999).

Fourth, different individuals have the expertise needed to evaluate re-

search in different fields. Finding all the needed expertise for comparing different fields in the same individuals is the exception rather than the rule.

Finally, there is the question of whose judgment should be final in resolving disagreements. This is a political question about relationships between and among the principals and the agents engaged in setting research priorities and performing the funded research. The primary principals are congressional and executive decision makers. Agency officials serve in a dual, intermediate role: they are the agents of the elected and appointed public officials, but to those who turn to them for funding, they are principals. The research performers, in the main, are the agents, but their participation in advisory councils and review committees also provides them with some of the influence and decision-making power more conventionally ascribed to principals. The very complexity of these arrangements complicates answers to questions about the distribution of power and influence among those who might decide and the distribution of benefits to those who might benefit. We return to this issue later in the chapter.

BRIEF HISTORY OF FEDERAL SCIENCE PRIORITY SETTING

Since the end of World War II, the salience of the issues of priority setting and retrospective assessment in U.S. science policy has waxed and waned. Priority setting received a surge of analytical and policy attention in the early 1960s, a period of steady increases of federal government support for research and development in the defense and nondefense sectors (Smith, 1990). The conventional benchmark of science assessment in this era is Alvin M. Weinberg's two articles entitled "Criteria for Scientific Choice" (Weinberg, 1963, 1964) and the surrounding exegesis (Toulmin, 1964; Smith, 1982). Weinberg defined a short list of generic criteria (see Table 3-1). Throughout the 1960s and 1970s, periodic efforts were undertaken to apply these and related criteria to priority setting in specific scientific fields (e.g.,

TABLE 3-1 Generic Criteria

External Criteria	Internal Criteria (How well is the science done?)	
	Is the field ready for exploitation?	Are the scientists in the field really competent?
Technological Merit		
Scientific Merit		
Social Merit		

SOURCE: Weinberg (1963).

National Academy of Sciences, 1965; National Research Council, 1972). Priority setting, framed in terms of the long-term prospects for science, also drew episodic attention from the U.S. Congress, the National Science Foundation (NSF), and the National Academies (e.g., Committee on Science and Technology, 1982; Irvine and Martin, 1984). The topic has continued to surface intermittently in the larger discourse on the federal science budget, such as in the report, *Allocating Federal Funds for Science and Technology* (National Research Council, 1995a) and in prescriptions by senior national science policy officials (e.g., Bromley, 2003). Priority setting is also implicit in the strategic planning undertakings of federal agencies, in which selected fields are chosen for emphasis, with explicit or implicit decisions made not to fund other areas or to alter relative distributions of support among areas.

Declining Attention to Criteria for Priority Setting

Although attention to priority setting has remained a staple component of U.S. government science policy, systematic attention to the development of criteria for scientific choice waned after the 1970s, at least in the United States. In part, this was due to the difficulties of moving from general agreement about broad priority-setting principles to agreement about the specific methods and measures to be used to operationalize these principles. In particular, somewhat in contrast to developments in European countries, where bibliometrics is often used in assessing and formulating science policy, there has been little consensus in the United States about the reliability or validity of techniques to assess the relative importance of different fields (Hicks et al., 2004). The decline in interest in applying systematic criteria for setting research priorities also partly reflected the unwillingness or inability of many scientific communities to agree about the priorities in their fields, as in the case of divisions in the high-energy physics community surrounding decisions to construct the superconducting supercollider. In addition, the few formal entities that existed to conduct prospective and retrospective analysis and to support research in concept development and tool building had short life spans.¹

General acceptance developed about relationships between the federal government as sponsor of basic and applied scientific research and the performers of this research. Congress and the administration, primarily through the legislative and budgetary processes, would set overarching national priorities for science and technology (although increasingly earmarking areas of research and performers); responsibility for converting these priorities into specific program areas and proposals devolved to intra-agency procedures and the peer or merit review system (Guston and Keniston, 1994; Guston, 2000). There was also general acceptance of the desirability of maintaining balance among fields in support for science, especially between

the life sciences and the natural sciences and engineering. However, few efforts were made for deeper analysis that might provide justification for important policy choices, such the relative apportionment of funds among disciplines and fields.²

For many years, far greater attention was paid to overall levels of federal government support of science than to questions of allocations among fields, reflecting a general level of satisfaction that for all its possible inefficiencies in terms of the goal of maximizing rates of return to public (or program) investments, the system had led to the U.S.'s preeminent position in world science. As Bruce Smith noted in congressional testimony in 1982, commenting on the National Academy of Sciences (1965) *Basic Research* report, "A common theme sounded by most of the panelists . . . was that the system we have evolved to support science, whatever our understanding of its inner mechanism, has given the United States a pre-eminence in the scientific world. Drastic changes in the present system, therefore, should be viewed with suspicion. "The quest should be for marginal adjustments in present policies to assure a continued United States leadership in basic science" (Smith, 1982:194). Similar analysis and recommendations pervade recent benchmarking assessments of the U.S. position in selected fields of science. As observed by the National Academies panel convened to assess the U.S. position in immunology research (Committee on Science, Engineering, and Public Policy, 1999b:52), the United States is the world leader in most major subcategories of immunology research, with this position being attributed to a system "that is largely an investigator-initiated, peer-reviewed, and merit-based system of awarding grants."

Thus, analytical and policy attention and research shifted in recent years to questions relating to the measurement of the social and private rates of return from research in general. This research has been conducted mainly within an economic paradigm that draws on earlier analyses pioneered by Nelson (1959) and Arrow (1962), which focused on whether competitive market dynamics could be projected to produce the socially optimal level of private-sector investment in research, especially basic research (e.g., Hall, 1996). The analysis derived from this framework, coupled with several major empirical studies on private and social rates of return to research in agriculture, health, and technological innovations, provided empirical support for the conclusion that governmental support of fundamental research yielded net social benefits. As a by-product, this research produced findings about benefit-cost ratios of investments in different lines of research, such as on different agricultural commodities (Evenson et al., 1979) and on diseases (e.g., Institute of Medicine, 1998; Gross et al., 1999; Murphy and Topel, 1999) that could have served as guides to future budget allocations among different directions of scientific research. Little evidence exists, however, that these findings in fact affected congressional budgetary allocations (Olsen

and Levy, 2004). Moreover, the findings tended to be directed at comparing target uses of research, not at comparative lines of research, the question posed by BSR.

Increasing Pressures for Systematic Priority Setting

The question of priorities among fields of science resurfaced in the 1990s and has gained increasing salience since then, as a result of several factors. First, the size and continued growth of federal research and development expenditures, together with increased competition for federal budget dollars, began to call forth new demands for accountability and demonstrated accomplishments. The enactment in 1993 of the Government Performance and Results Act (GPRA) focused those demands: it required that all federal agencies develop multiyear strategic plans and evaluate and report annually on their activities in relation to the objectives stated in these plans. For research agencies, GPRA created pressure to implement systematic methods and bureaucratic routines for assessing the value of research investments (National Research Council, 1999, 2001c). The act shifted attention from statements of an agency's needs and opportunities toward outputs and outcomes. Inexorably coupled with demands for accountability were new demands coming from both the administration and Congress for "evidence" and documentation of performance and results. The demands are manifest in GPRA and in its implementation via the Performance Assessment Rating Tool (PART), created by the Office of Management and Budget (OMB). They continue to receive support from congressional and administration leaders.³

PART was introduced to cover selected federal agencies as part of the FY 2004 federal budget process and has since been applied to an increasing number of agencies and programs. As described on the OMB web site, "PART was developed to assess and improve program performance so that the Federal government can achieve better results. A PART review helps identify a program's strengths and weaknesses to inform funding and management decisions aimed at making the program more effective. The PART therefore looks at all factors that affect and reflect program performance including program purpose and design; performance measurement, evaluations, and strategic planning; program management; and program results. Because the PART includes a consistent series of analytical questions, it allows programs to show improvements over time, and allows comparisons between similar programs" (available: <http://www.whitehouse.gov/omb/part>).

A second factor has been the significant advances that have been made in theoretical and especially empirical studies of scientific activities since the Weinberg articles of the 1960s. Significant advances have been made

in assembling and making more accessible quantitative data on several aspects of scientific activity that previously required labor-intensive effort and that were difficult to link together. For example, there has been continuing expansion and refinement of the NSF's Science and Engineering Indicators biennial reports, leading to more readily accessible data on publications, patents, and patterns of collaboration among scientists.

Data on scientific activity, including numbers of publications by keyword, numbers of citations, and so forth, are now readily available online through such services as Thomson Scientific's Web of Science® and Google's Google Scholar. The National Bureau of Economic Research has compiled an extensive data set of U.S. patents that includes all citations to these patents and a broad match of these patents to financial data sets (Jaffe and Trajtenberg, 2002). At least two major handbooks have been published, distilling a much larger and diverse literature on quantitative methods in the use of publication and patents statistics in studies of science and technology systems (van Raan, 1988a; Moed et al., 2004). New methodological and empirical ferment is emerging in the use of network theory and career trajectories to explore patterns of collaboration, and thus leader-follower relationships and the diffusion of ideas and techniques, among scientists (Wagner and Leydesdorff, 2005). Advances in data mining and data visualization techniques facilitate processing large quantities of data, making it possible to identify or confirm relationships that may previously have been unrecognized (e.g., Boyack and Börner, 2003).

Advances in quantitative measurement and data analysis have also facilitated the development of new theories of the relationships among scientific activities and their effects in the larger society, and increasingly sophisticated theoretical and empirical models have been tested for examining these causal relationships. Thus, for example, researchers have examined linkages between bibliometric data (on scientific publications and citations) and patent data to advance conclusions about the productivity of federal government investments in some areas of basic research (Narin et al., 1997). Advances in data availability and analysis make the quantitative assessment of developments in a nation's scientific enterprise increasingly feasible and attractive to public-sector officials. With all these advances, the prospect of using quantitative analysis systematically to channel public funds to their most productive scientific uses appears more attainable than before.

A third factor making priority setting more salient involves the dynamics of science itself, especially the widespread consensus that the greatest opportunities for advances in science now involve the crossing of traditional disciplinary boundaries and the creation of new fields. NIH's Roadmap, for example, addresses what agency leaders describe as revolutionary and rapid changes in science and the need to overcome barriers created by the complexity of NIH as an institution comprised of many units; the compart-

mentalized structure of the NIH bureaucracy, with its division by organ, life stage, disease, and scientific discipline; and the rapid convergence of science. Related to this judgment are increasingly voiced concerns that the combination of existing organizational arrangements and procedures that science agencies, including NIH, use for setting priorities, selecting research proposals, and evaluating the outcomes of research, together with increased competition for funding, are leading to unduly conservative, risk-averse selection of research awards.⁴

Finally, the globalization of scientific activities, coupled with the widespread belief that scientific leadership is increasingly linked to international economic competitiveness, introduced another issue into science policy discussions. A report from the National Academies (National Research Council, 1995a), articulated the principle that a country's position relative to its scientific competitors should be taken into account in deciding the distribution of resources among scientific fields: "The President and the Congress should ensure that the [federal science and technology] budget is sufficient to allow the United States to achieve preeminence in a select number of fields and to perform at a world-class level in the other major fields" (p. 14).⁵

All these forces have given impetus to recent efforts to formalize science policy decision-making processes. Almost reflexively, there have been increasing calls for quantification and for transforming the more extensive data on science and technology and improved techniques for the analysis of such data into science metrics (for inventories of widely used metrics, see Geisler, 2000; National Science Board, 2004).

DEBATE OVER PRIORITY SETTING AND ASSESSMENT MECHANISMS

As already noted, peer review has long been the dominant approach in federal research agencies for evaluating the past performance and future potential of research areas and for setting priorities. Peer review is essentially a clinical and deliberative process and one that relies heavily on the expertise of working researchers. It is used most commonly to evaluate proposals for research projects or programs coming from single investigators or research groups; less commonly, the same approach is used to advise research managers on broader matters, such as evaluating the past performance or future prospects of entire research programs or selecting priorities for the future development of these programs. Whatever the purpose, the general approach is similar. Research managers convene groups of experts in relevant research fields, typically constituted as peer review panels, visiting committees, or advisory boards, that deliberate on issues or choices presented by research managers. They are sometimes informed by

additional input, for example, reviews solicited from specialists in particular narrow areas being considered. In making their recommendations to higher level agency decision makers, research managers draw on the judgments of these deliberative groups and add in their own judgments to the extent their agency prescribes or allows. Devolution of decision-making authority, or in this case, recommendations, to peer review panels is the “special mechanism” by which the social contract for science “balances responsibilities between government and science” and thus fosters accountability (Guston and Keniston, 1994:8).⁶

Outside reviews of research agencies’ efforts to assess programs and set priorities have generally endorsed the clinical, deliberative methods of expert review as the best way to assess research fields. For example, a National Research Council study committee that reviewed agency experiences under GPRA concluded: “The most effective way to evaluate recent programs is by expert review. The most commonly used form of expert review of quality is peer review. This operates on the premise that the people best qualified to judge the quality of research are experts in the field of research. This premise prevails across the research spectrum, from basic research to applied research” (National Research Council, 1999:39).

Accountability Challenges to Peer Review

However well peer review as a method of research assessment may have served science agencies, the scientific community, and society in the past, this approach has recently come under challenge. A major challenge has come from the movement toward greater accountability and attention to performance management, as embodied in GPRA and PART. GPRA requires federal agency managers to establish strategic goals and to demonstrate that these goals have been met, on the basis of predefined measures of performance. Development of standardized outcome measures is also seen as a means by which science managers can compare the bang for the buck across different kinds of expenditures. PART, as already noted, requires an increasing proportion of federal programs to apply a consistent, evidence-based approach to performance measurement, extending from the specification of strategic goals (such as lives saved) to outcomes and outputs. Development of performance measures is encouraged because such measures are seen as the ultimate results for the public. But the PART guidelines also note that “the key to assessing program effectiveness is measuring the right things,” by which is meant, “measures that meaningfully reflect the mission of the program, not merely ones for which there are data” (p. 16).

Intended to be broadly applicable across all federal programs, the PART procedures also contain a specific set of criteria to assess the effectiveness of the federal investment in research and development. The three salient

criteria applicable to NIH-NIA programs are relevance, quality, and performance. For each criterion, the use or development of quantitative metrics is emphasized. Thus, in considering relevance, the PART document states that “OMB will work with some programs to identify quantitative metrics to estimate and compare potential benefits across programs with similar goals. Such comparisons may be within an agency or among agencies” (p. 56). GPRA and PART are similar in that both provide agencies with pressures or incentives to move toward more quantitative methods for setting priorities or assessing performance.⁷

Despite the skepticism that researchers have at times expressed about applying quantitative approaches to assessment of their work—for example, concerns about the spawning of “LPUs” (least publishable units)—there are valid reasons for trying to use them. An important one is that many research agencies’ expenditures are justified not only in terms of advances in pure knowledge but also in terms of their potential value to society, some of which are eminently quantifiable. Among these, depending on the agency, are improved health or longevity, education, environmental quality, and public safety and security. Already, there are efforts under way to develop measures related to the impacts of research on some societal goals, with current emphasis being on improving the reliability, timeliness, and administrative feasibility of the measures (U.S. Department of Energy, Office of Science, 2004).

The prominence of the “health and well-being of older Americans” among the strategic goals of NIA makes it tempting to quantify at least those outcomes and to seek evidence for causal links between research and those societal benefits. Quantitative measures of many of the benefits, both realized and projected, as well as benefit-cost ratios, already exist. For example, healthy 70-year-olds live longer and spend less on lifetime health care than their less healthy peers. In one study, individuals with no functional limitations had a life expectancy of 14.3 years and expected cumulative health care expenditures of \$136,000 in 1998 dollars, while those with one functional limitation had a shorter life expectancy (11.6 years), but could expect to spend more on health care (\$145,000) (Lubitz et al., 2003). If health promotion efforts (e.g., exercise, smoking cessation) can improve the functioning of older Americans, these benefits can be predicted to follow. Similarly, observed benefits of cognitive and affective phenomena for health (e.g., Rosenkranz et al., 2003; Levy, 2003) might also be quantified in economic and life expectancy terms. The estimate that a 1 percent permanent reduction in mortality would be worth about \$500 billion (Murphy and Topel, 1999) makes possible benefit-cost analysis of investments in mortality reduction.⁸

Research agencies’ internal needs also create pressure for quantifying research progress. They need to compete successfully for research funds

with other uses of federal funds that are justified in accounting terms in the tightening discretionary budget for nondefense science and in the face of competition from nonresearch priorities. And they need to do so at a time when demand for research funds is increasing as a result of technological advances, methodological developments, and growing concern with complex systems and interdisciplinary problems that require more expensive capital-intensive and team-based research. Some new research areas require major investments in expensive technology.⁹ Public officials want a rational basis for making difficult choices, and quantitative measures are attractive because they can be defended as “objective.” Some research agencies have also identified internal reasons to seek quantitative measures of research performance. For example, such measures might be useful for justifying budgets against the claims of other units in the same government department and for resisting pressure to shift expenditures in ways that would benefit specific political constituencies at the expense of the scientific and practical benefits of research to the public (see, for example, National Research Council, 2005c).

Challenges to Quantification

Reliable and valid quantification of benefits from scientific research would obviously be desirable for assessing the value of past investments in research. Such output measures would provide research managers and higher level government officials with valuable yardsticks for evaluating past investments and a counterweight to inappropriate claims on research budgets from interested groups. However, developing reliable and valid performance measures that work across disparate fields has been very difficult. Questions continue to be raised by policy makers, research administrators, practicing scientists, and specialists in program evaluation about the reliability and validity of the basic data series; about errors in measurement; about the ability of actors in the scientific enterprise to manipulate or “game” several mainstream quantitative techniques, for example, by pooling citations; and about the applicability of techniques used to study the workings of the scientific enterprise to evaluation and priority setting (e.g., van Raan, 2005; Weingart, 2005; Monastersky, 2005).¹⁰ Particular quantitative methods have also been criticized. For example, benefit-cost analysis has frequently been criticized on the grounds that many of the costs and benefits, especially the latter, are not traded in markets and therefore require inferences and imputations before seemingly precise quantitative calculations of value can be made (e.g., Gramlich, 1981; Stiglitz, 1988). Similarly, knowledgeable observers have raised the concern that the use of performance scorecards as represented by PART-like mechanisms is becoming detrimental to the conduct of science (e.g., Perrin, 1998; Weingart, 2005).

Critics of the use of performance measurement in social and health policy further argue that these policy goals are not like market-oriented or industrial production, in which output measures, such as numbers of items produced or the value of sales, are appropriate and readily linked to production activities (see, e.g., Hatry, 1989). They believe it is much more difficult to measure how much schoolteaching contributes to student learning or how much scientific research programs contribute to new socially beneficial discoveries (see Cozzens, 1997).

Reservations have also been advanced regarding the construct validity of performance measures as applied to scientific endeavors. Many scientists and science managers generically reject the idea that scientific progress can be measured in terms of discrete, homogeneous outputs, analogous to number of miles of road paved or speed with which social security checks are processed. They contend that scientific advance is inherently an uncertain process that often takes or even requires an elongated, circuitous path. Important advances often appear unexpectedly and from unlikely sources; long time lags may occur between a scientific development and its application; findings are used in ways not conceived of either by researcher or sponsor; findings deemed interesting but not significant take on new import when combined with newer findings or applied to newly emerging situations.¹¹

Critics of quantification also challenge the value of retrospective assessments for research priority setting. Although past performance is often seen as the best predictor of future performance for individual researchers, the recent performance of a research field may or may not be a good predictor of whether additional investment in that field is likely to lead to great advances or to less productive elaboration of past work. Of special concern is that the predictive value of the past for the present may well decline at scientific turning points, when discontinuous leaps or falls occur in the scientific richness of a new or established field. As one scholar noted, "Although every scientist is aware of impending revolutions, no clear universal sign tells even the most astute observer the area of science in which the next revolution will occur or what form it will take. The most brilliant scientists are not able to predict exactly the kind of revolution they themselves will be making" (Cohen, 1985:21).

Critics also argue that valid measures are hard to devise and defend for scientific research within the output, outcome, and impact frameworks of GPRA and PART because research in a single area may yield several kinds of outputs and because each research product may produce several different kinds of value (outcomes). Also, the impacts may occur so far into the future and may require so many complementary activities that extend beyond the influence of either the scientist or program manager that considerable patience and carefully crafted analysis are necessary to establish or refute

causal links from funding in specific fields to specific societal impacts (David, 1994; Radin, 2000).

Performance measurement applied at the agency, division, or program level is also criticized as setting up competition of the parts against the whole, with each budget unit seeking to claim credit for or internalize all the benefits associated with its activities, rather than participating in activities that may produce larger but more widely dispersed benefits. As phrased by David (1994:297-298), “we need not move in the direction of taking apart the very complicated system of science and technology research, which works in ways that not all of us fully understand, and making each of the bits of it compete with one another in the claims they make for the performance of the system as a whole. To point toward the larger outcome goals, which could be imputed at the systemic level, and to try to get people to lay claim to bits and pieces of that one, or their contribution to that other one, is probably the wrong direction in which to push the formation of science policy thinking.”

In addition to these points, many of the outcomes of research, such as satisfying curiosity about the universe or understanding human society, are intangible and hard to put on a scale that permits comparison with other kinds of returns to research or to nonresearch expenditures. To further complicate the measurement problem, the nature of the products and the kinds of value they may produce are often unknown in advance. For all the past and ongoing efforts invested in developing improved forecasting or foresight techniques, predictions about societal impacts of scientific and technological advance are viewed with good reason as highly speculative (for the emerging case of nanotechnology, see Roco and Bainbridge, 2003).

It is also argued that assessment criteria, even if valid for gross discriminations about the routine progress of science, are much less useful when applied to discrimination in the tails of a distribution. Scientific innovation is heavily concentrated in the far upper tail of accomplishment in science: thus, criteria that are effective in discriminating reasonably good from reasonably bad normal science are likely to be unproductive or even counterproductive in predicting events, trends, or productivity in the upper tail of the relevant distribution, where breakthroughs occur.

Yet another reservation about current efforts to quantify the performance of research investments is that few agencies systematically treat the development of human capital as an output complementary to conventionally measured research outputs. In one view, however, “the most important contribution that is being made through basic research funding to national economic growth comes not through the transfer of research findings directly but through the transfer of knowledge and skills of trained personnel who move from the university laboratory into employment in the private sector” (David, 1994:297).

Quantitative assessment of science carries a substantial risk that researchers will behave like teachers “teaching to the test,” using an analogy to one response to quantification in educational measurement. For example, if progress is measured by number of publications per dollar, measurement creates incentives to produce the smallest publishable units and against work with long gestation periods leading to major breakthroughs (Butler, 2004).

Perhaps the most seriously questionable use of quantitative measures of research output is for making direct comparisons of research productivity among research fields. Such comparisons, however, are central to the questions posed by BSR. Different fields produce different kinds of output, even at the level of readily measured products, such as journal counts. Lags between submission of manuscripts and publication vary across both journals and disciplines. Conventions concerning references to existing literatures also vary across fields, affecting the frequency with which specific articles may be cited by relevant scientific communities.

The problem of accurately accounting for output, productivity, and impact is compounded when differences in publication outlets are considered. For example, researchers in some fields mainly publish in journals indexed in major databases, while researchers in other fields mainly publish books, monographs, or technical reports that are not so indexed. This is particularly true in the social sciences, which Hicks (2004) has described as having four literatures—international journals, books, national journals, and the nonscholarly press. To the extent that different disciplines in a program manager’s portfolio have different publication patterns among these four literatures, quantitative measures based on bibliographic, journal-centered databases may be biased indicators for comparing scientific output.

Identifying these concerns is not equivalent to ruling out the utility of quantitative methods, including bibliometric data ones. For several of the examples cited, systematic collection of data, say on average delays in the submission-acceptance-publication process, would provide a ready means of adjusting data to move toward “unbiased” estimators. The more pressing concerns in those examples relate to (1) the need to be aware of the limitations of raw or unadjusted measures; (2) the considerable technical sophistication at times required to adjust or refine the measures; (3) the added time and costs involved in making the necessary adjustments so that the data are available and comprehensible in time frames consistent with agency priority and budget setting processes; and (4) the likelihood that the proper interpretation even of carefully adjusted measures may remain a matter for legitimate dispute.

Beyond issues associated with aspects of performance measurement, comparison across fields of inquiry is complicated by the different kinds of questions or problems and the different patterns or levels of effort that are

required for analyses in different research areas. Some areas present well-defined, delimited problems that call for a concentrated effort to obtain a solution, and then they go away. An example is research on the impact of Medicaid spend-down rules on the asset management and housing decisions of aging families (e.g., Adams et al., 1992); the results of that research could be incorporated into policy, and there was no need for further research. Other areas present core questions that require continuing attention. These include the reliability of self-reported health status in surveys for predicting health needs and outcomes and the development of data infrastructures (e.g., the Health and Retirement Survey) on which scientific investigations can build. Finally, there are areas in which the solution of a well-defined problem leads to new questions or methods for research.

The criteria used to judge scientific fields must be sufficiently nuanced to recognize these different paths of progress. For well-defined problems, when has scientific analysis definitively succeeded or failed, justifying termination? When has the end of a successful research effort opened new opportunities that deserve increased support? For continuing research issues, which scientific approaches are fresh and promising, and which are stale?

The situation is the same for other quantitative indicators. Research in some fields leads to patentable inventions, while in others it may lead to improved practices or new policies.¹² Research in some fields leads to new drugs or medical procedures, whereas research in other fields leads to less readily quantifiable medical benefits, such as improved diagnostic categories or ways of interpreting diagnostic tests. Behavioral science research may lead to valuable advice to individuals about ways to change their behavior. However, when no organization has a strong incentive to publicize this advice, behavioral change may not be a fair test of the value of the science.

The challenge of commensurability—“the comparison of different entities according to a common metric” (Espeland and Stevens, 1998:313)—of outputs and outcomes across fields would appear to be particularly formidable in areas of social and behavioral science because the outcomes resulting from such research typically take the form of new knowledge that might be applied in practices and policies, rather than tangible objects. Whether or not new knowledge is used, and how it is used, depend on a variety of factors in addition to knowledge production itself (e.g., Weiss, 1979; Landry et al., 2003). Knowledge may lead in directly observable and traceable ways to changes in practice or policy, but the effects are more often indirect. For instance, knowledge may expand or alter the set and value of options considered by decision makers—an “enlightenment value” of knowledge (Weiss, 1979) that is important even when the specific action taken is unchanged. Thus, there is no straightforward link from knowledge production to its application, an application or its absence cannot be attributed in any simple way to the actions of scientists or research manag-

ers. For many research allocation decisions, the question before program managers or advisory groups is not readily amenable to a cost-effectiveness framework that would allow for quantitative analysis, even with specified uncertainty bounds, of which line of inquiry holds the highest promise of yielding satisfactory answers. If different fields address different questions and produce answers that lead to different kinds of outputs, the first and continuing challenge is first to agree on a standard unit of measurement.

These methodological critiques of quantification of scientific results complement the concern among many scientists that quantification *à la* the PART procedures is a threat to the traditional primacy of expert peer review. This concern is rooted in the above methodological concerns, in the idea that the judgments that emerge from expert review panels provide a more thoughtful and nuanced assessment of scientific progress than can come from any available quantitative methods, and in a concern that quantification entails a shift of power and influence over priority setting from working scientists to government officials following bureaucratic procedures. This last concern, to adapt Oscar Wilde's comment, is about the possible ascendancy of nonscientists who know the price of everything and the value of nothing.

Alternatives to Quantification

Because of the limitations of quantitative indicators of scientific progress, some agencies have sought to justify nonquantitative methods as responsive to accountability needs. For instance, the NSF has received permission from OMB to employ qualitative methods for assessing level of performance set against agency strategic goals (National Science Foundation Advisory Committee for GPRA Performance Assessment, 2004). The NSF performance reporting system employs expert judgment via panel and mail reviews to vet proposals, augmented by a system for periodic review of program-level activities by committees of visitors. The NSF approach to assessment has relied on internal documents, committee of visitors' reports, and a database of accomplishments, among other sources of information.

Other research funding organizations, including NIA's BSR Program, have also seen considerable merit in preparing similar collections of information as inputs to expert judgment. For example, BSR has prepared numerous narrative descriptions of "research highlights," "science advances," and "stories of discovery" to document the results from the research it funds (e.g., Behavioral and Social Research Program, 2004). These brief histories highlight the contributions of agency-supported research to improvements in health or well-being and demonstrate by example the value of the entire program. A historical approach also allows readers to appreciate the different kinds of value that result from different lines of research, even though

it does not attempt quantitative comparisons. For example, one of these narratives describes the role of program-funded research in showing that increased longevity during the 20th century has not resulted in longer periods of disability. Another shows how funded research has provided better data for understanding the relationships between socioeconomic status and health. Yet another demonstrates that job control at work is a major risk factor for cardiovascular disease in men. BSR combines these narratives with a variety of outcome indicators, such as lists of peer-reviewed publications, honors received by funded researchers, and recognition of funded research in the specialist and popular press, to help inform expert judgments by its advisory board.

Case histories have not always proved useful, however. Earlier studies, such as the 1968 Technology in Retrospective and Critical Events in Science (TRACES) study (Illinois Institute of Technology, 1968), sponsored by the NSF, and Project Hindsight, sponsored by the Department of Defense (Sherwin and Isenson, 1966), which sought to relate advances in fundamental science to important technological advances, proved to be expensive and to have limited persuasive impact. Nettlesome methodological disagreements arose about the validity of the findings (Kreilkamp, 1971), and few such large-scale endeavors have been undertaken in recent years.

Differences of opinion also exist among program managers regarding the external validity and political impact of historical accounts or case studies. They can be subject to selection bias, especially when selected by program managers who have reasons to show a program's best face. Thus, some argue that without large-scale comparative studies, historical accounts lack persuasiveness about the actual contributions of a program and are unlikely to convince higher levels of management. To others, however, one compelling case can be akin to the picture that says a thousand words. In addition, the case history technique in general fails to adequately satisfy OMB expectations that agencies install data-based management systems to monitor performance.

The current situation is thus characterized by both methodological and policy turmoil and disagreement. Some agencies seek to develop, validate, and apply quantitative measures of research output and its value. Others see available quantitative metrics as hopelessly inadequate for their assessment purposes and believe that expert judgment is the only valid and appropriate way to evaluate the past performance or future potential of research (see National Research Council, 1999). Nevertheless, the trend toward increased quantification is clear. A combination of forces—OMB mandates, improved sophistication in quantitative methods, more critical examination of the limits to which specific methods can legitimately be pushed, more modest claims on the part of advocates and practitioners of quantification, and what may best be termed resignation to the use of quantitative

methods—has softened the edges of earlier either-or debates about the use of expert judgment/peer review procedures versus quantitative methods for priority setting and assessment.¹³ Consensus may be emerging about the need for informed expert opinion based on the “proper” use of quantitative methods by relevant experts.

Limits of Expert Judgment for Comparing Fields

When research managers must set priorities across research fields, many of the problems of comparison faced in quantitative assessment also apply to expert judgment. Panels of experts that may be able to give reliable advice in specific, long-established disciplines may have much greater difficulty advising across fields. In a well-defined scientific discipline or field, it is reasonable to presume that the appropriate standards of judgment are well understood by most of the experts who might serve on a peer review or advisory panel, even if the standards are not identical in all parts of the field. The members of such groups understand each other’s work, and it would be possible for other experts in the same field to assess a panel’s findings by applying the same standards. It is less safe to presume that such shared understanding and the attendant possibility for checking judgments exists when review panels are organized across more varied areas of expertise. The problem is likely to get worse, the greater the breadth of the set of programs or units that are being compared.

Suppose an agency empanels a broadly multidisciplinary expert group to evaluate aspects of a multidisciplinary program. Group members must either judge outside their expertise or rely on their colleagues’ judgments, in which case they may fail to understand the standards that their colleagues are applying. Not being familiar with the content of the work being proposed or its potential for opening up new lines of research, experts sometimes use methodological rigor as a default evaluation criterion. The result may be, as has been increasingly asserted with regard to both NIH and NSF, that review panels are inherently too conservative about supporting radically new or transformative research ideas over well-crafted mainstream but incremental science.¹⁴ Strong criticism by one or two review panel members, particularly on specialized matters in those members’ areas of expertise, may be enough to defeat an idea. Similarly, panels have been criticized as favoring science in established disciplines over interdisciplinary proposals (National Research Council, 2005b:Chapter 6). It has been claimed that experts on a panel defer to the judgment of each member in his or her own field, with the result that fields perpetuate themselves even when their potential for generating important advances is weak and when a broader analysis of a research agency’s portfolio would justify reallocation of funds elsewhere.

The systemic logic behind these behaviors, according to Brenneis

(1994:31), is that expert review panels employ a “fairness through apparent clarity” model of decision making. In this model, scholarly progress is seen as incremental, so that proposals are favored when they “are clearly linked to a sense of how ‘science works.’” Proposals that promise to break new conceptual ground or to challenge and refigure dominant paradigms are viewed not so much as ‘bad’ proposals but as difficult to evaluate and compare with other contenders” (see also Guetzkow et al., 2004).

Collective judgments by peer review panels may thus not have a clear meaning, especially when a panel is covering a multidisciplinary range of fields. The problems cannot be solved by combining judgments from different disciplinary groups because standards in different fields may not be comparable. Assume that there are established fields of science, characterized by the conventional attributes of disciplines; that is, journals, professional associations, academic standing as departments, and institutionalized legitimacy within a federal science agency, such as an established study panel or a directorate or division devoted to the support of each field.

Assume further that over time one field becomes insular; that is, it focuses on problems that engage specialists but do not look important from an outside perspective, either on scientific or practical grounds, and that prove unproductive in retrospect. The experts in such fields nevertheless continue to believe that they and their colleagues are engaged in exciting, productive, and societally relevant work. When asked to judge recent or proposed new work on such criteria as originality, they rate the studies as more original than they would appear to outsiders to the insular field. Researchers in such a field can point to a steady stream of output, say, in articles published in leading journals in the field and citations to these articles, albeit predominately by other researchers in the field. Without some external yardstick, it is not possible to know whether or not a particular field is such an insular and moribund field in which the collective judgment of its experts is untrustworthy. Research managers want to identify such fields sooner rather than later, but the judgment of experts from within the field may be misleading, and the judgments of multidisciplinary peer review groups may also fail to offer good guidance.

The problem of identifying fields that have passed their prime is at the core of the questions posed by NIA-BSR. Its concern is that it may be supporting some fields that are producing minor if technically well done advances in knowledge derived from long established but increasingly stale paradigms, while choosing not to fund other fields, theories, methods, and findings that promise (with uncertain likelihood) to yield significant new advances in fundamental knowledge that will illuminate not only the field from which they come, but also spill over to enrich other fields or even create new ones. Although peer review panels in many fields believe that there is much high-quality work in those fields, it is possible that the experts in

some fields have an inflated view of the value of research in those fields. It is out of such concerns that research managers in BSR seek for a trustworthy method or strategy of research assessment that would make it possible to evaluate the hypothesis that there are serious imbalances in the value of research across the fields funded by the program.

As noted throughout this report, these concerns have long antecedents. But despite various efforts at broad-scale retrospective and prospective judgments (e.g., Deutsch et al., 1971; Irvine and Martin, 1984; Inkeles, 1986; Abrams, 1991; Henkel, 1999), these questions are generally finessed. One reason is their extreme sensitivity in scientific circles, because any explicit effort to compare the value of research across fields entails making invidious distinctions, with possible damage done to the field(s) given lower evaluations in future funding cycles.¹⁵ In contemporary U.S. science policy discussions, tactful discourse centers on the concept of balance and may suggest that one or more areas of science are underfunded, but not that the questions being addressed or methods being employed by other fields are in any manner experiencing flagging vitality.

The current budgetary pressures and the additional impetus for accountability and quantitative measurement of research progress make it increasingly difficult to finesse the questions of comparative assessment. Despite the difficulties of comparing the research efforts of different fields on the same quantitative scale, demands for accountability create serious pressure to provide clearly stated rationales for recommendations and decisions about priorities among research fields.

RESEARCH ASSESSMENT AND THE ISSUE OF POWER

As already noted, a critically important but sometimes unacknowledged issue behind debates about how to assess science is that the choice of method may both reflect and affect who has power and influence. Particularly at issue here may be the relative power and influence among researchers, research managers, and nonscientists in government. Such issues have been present in science policy since the beginnings of sponsored research in the United States, when the sponsors were private foundations, such as the Carnegie and Rockefeller Foundation groups (see Box 3-1). The issue continues into the present. An increased emphasis on the use of quantitative indicators in science policy decisions, especially to the extent that indicators can be developed by technicians who are not researchers in the relevant fields, can easily weaken the influence of scientists vis-à-vis agency science managers, or of scientists in general vis-à-vis nonscientist decision makers in government.

Bibliometrics provides a good example of the issue of power, latent in many current discussions of the use of “objective” measures of research

BOX 3-1

Power Relations in Sponsored Research, 1900-1945

In the United States, the system of sponsored research grants evolved in private foundations, especially those of the Carnegie and Rockefeller groups. At first, the idea of programmatic, actively managed grants to individuals met resistance because it seemed to transgress entrenched individualistic values and the belief that scientific discoveries are acts of individual genius. In the 1920s, foundations evaded these problems by giving block grants to universities or research institutes with no strings attached: they were designed to develop entire communities, not advance particular individuals' lines of work. But these programs proved unaffordable after 1929, and in the late 1930s, the Rockefeller Foundation pioneered a system of programmatic grants to scientists in a few fields that were deemed by program managers to be of strategic value.

A new social role of grants manager evolved. Foundation officers earmarked strategic fields for investment and, in the absence of a system of peer review, selected among applications for support. Grant managers became partners in science, in direct and unmediated relations with their grantees.

This active relationship profoundly upset existing power relations with foundation boards of trustees. Trustees, who were mostly practical men of business and who had previously made decisions on the basis of their ability to judge organizations, were now in a position of rubber-stamping decisions made on technical grounds by mid-level managers. They were effectively deskilled as experts in organizing productive labor. Senior scientists on boards took the same line, opposing "planning" in science, even though relations between grantees and program managers were remarkably untroubled. In practice, activist managers were helpful, not intrusive, for example, in foster-

progress. Resistance to the use of bibliometric measures by academic researchers surfaced almost immediately upon their introduction. "The reaction was predictable," according to Weingart (2005:118) "because first of all the very attempt to measure research performance by 'outsiders,' i.e., non-experts in the field under study, conflicted with the firmly established wisdom that only the experts themselves were in the position to judge the quality and relevance of research and the appropriate mechanism to achieve that, namely peer review, was functioning properly." Adopting any method of assessing research potentially affects who has the ability to influence the setting of broad research priorities, the contours of specific programs with respect to subfields and methodologies, and decisions concerning which

ing communication among scientists in different disciplines. Grantees quickly realized that.

Trustees distrusted program managers because they could not see what would prevent them from abusing their new power to set priorities and decide on individual proposals. Contention between managers and foundation boards changed only when managers, such as Warren Weaver at Rockefeller, showed their ability to devise and manage programs of individual grants in selected fields like genetics or molecular biology, a field that Weaver helped to define.

The system that evolved by the late 1930s had trustees appointing program officers, who were empowered to select strategic areas for investment and to actively manage systems of individual grants. Scientists accepted the active participation of program managers in directing research along selected lines, but they retained complete control of how the actual work would be done. Once grants were made, program officers never interfered and declined invitations to advise on the particulars. This division of labor worked without advisory committees, peer review panels, or formal procedures of reporting and accountability.

Communication was the main reason the system worked. Program officers worked constantly to be well informed of trends in the fields they sponsored and in the activities and reputations of leading figures in the fields. They did this by identifying a few trusted individuals who they learned would provide objective and disinterested advice and by continual traveling and conversing informally with grantees and potential grantees, including younger scientists. Program managers could become effective partners in science because they understood the personalities and intellectual politics, as well as the science, of their areas of interest almost as well as the insiders did themselves.

SOURCE: Kohler (1987, 1991).

proposals to fund (and at what levels) and which to reject. To the extent that the critical information needed for making research portfolio decisions can be gained without reliance on the researchers themselves, the power to make those decisions can be shifted from researchers to research managers. In short, decisions about quantification of scientific progress have a power dimension, whether or not this is within the awareness of those involved, and a shift in power relations can have significant consequences for the directions of science.

Such a shift may be viewed as good for science—for example, if research managers have a better overview than scientists of opportunities across many fields, or a better appreciation of which research directions are most likely

to meet societal needs or agency priorities. It may also be viewed as bad for science (for example, because of the possibility of political or ideological interference with scientific research agendas, as in the controversy over the claim that management at the National Endowment for the Humanities has overturned peer review endorsement of proposals because they address issues considered sensitive by political appointees; see Jaschik, 2006).

The issue of power is implicit in the implementation of GPRA and PART. To the extent that these tools emphasize routinized measurement of easily quantified attributes of research, they shift power away from the judgments of the scientific community and toward others, such as those who devise the indicators and those who can find ways to game the assessment system. Efforts to gain approval for assessment mechanisms that rely more on the judgments of scientists, apart from claims that they provide better quality assessments, are in part efforts to prevent a loss of influence by scientists over science priority setting.

It is important to recognize in this context that research managers at NIH are scientists as well as managers. They are typically in the job classification of Health Scientist Administrator, which is taken to mean that their first calling is as a scientist and that the administrator role is secondary. NIH program managers typically hold advanced degrees in science, not management. Once in their positions, however, research managers are expected to act as stewards of their scientific fields in the context of the mission of NIH and their institute or center. The challenge in such a position in any federal science agency is to stay abreast of one's scientific field: to know what are the emerging opportunities and challenges, who is doing outstanding work, who is coming up in the field, who is on the cutting edge, what the demands are in the field for technologies or models, and so on.

In a stylized manner, much of the research support provided by NIH, especially that occurring in the form of investigator-initiated (R01) projects, represents grassroots initiatives of independent individual researchers. In this model, the frontiers of science, both in terms of the questions (or puzzles) posed and the selection of projects to answer the questions, are determined mainly by the collective workings of the scientific community.

NIH program managers ideally function as part of this community, not only as scientists but also as advocates, stewards, and occasionally as entrepreneurs for research fields. A field may need a research tool that NIH can support and make available; better access to data; or a new way of organizing research. A prominent example from BSR is its support of the ongoing Health and Retirement Survey at the level of about \$10 million per year. This survey provides data useful to a large number of individual research projects. As noted in Chapter 2, research managers in NIA/BSR have been proactive as entrepreneurs of research by using the management tools at their disposal to provide such collective goods for science. Initiatives

by research administrators can be instrumental in starting or accelerating the development of a field or line of research, and research managers may need to have an entrepreneurial spirit to go along with their understanding of the science and the needs of a field. They may be called on to advocate for specific projects, to urge the adoption of policies, or to work with colleagues inside (or outside) NIH to build support for programs, grow the funding, or help further the appreciation of the science. They need to know how their field relates to others so they can effectively and enthusiastically cooperate in new, high-priority interdisciplinary areas. They also need to be alert to the scientific human resource development issues critical to the programs they administer.

It is possible that a research administrator can see opportunities or challenges that are as yet not fully visible to most working scientists in a field, and thus come to a judgment about a field's needs that differs from the consensus of those working in the field. Research managers' judgments may differ from those of active researchers because of the greater value the former group places on the relevance of research to an agency's mission. And administrators responsible for several fields often make judgments about priorities that differ from the consensus judgments in some of the fields. Differences in judgment may arise from differences between managers and working scientists in the weights assigned to different program objectives or because they use different methods of assessment.

Research managers acting as stewards and entrepreneurs may try to convince higher organizational levels, elected officials, and the scientific communities with whom they interact of the importance and relevance of ongoing or emerging fields of science. Research managers may be more or less entrepreneurial and more or less successful in this role, depending on personal disposition or agency structure, practice, and culture.

In the best case, working scientists and science managers can bring valuable and complementary perspectives to the task of assessing science. In designing methods for assessment and priority setting, then, it makes sense to avoid framing either-or choices between mechanical, quantitative, and bureaucratized decision making led by science managers and qualitatively informed, nuanced choices dominated by scientists. The proper questions to ask in guiding research assessment and priority setting do not concern whether to use quantitative measures, but what should be the appropriate roles of quantitative measures and of deliberative processes of peer review and how should the perspectives of scientists and science managers be combined to provide wise guidance for science policy decisions.

The above observations on the role of program managers are mediated by the formal structures and informal practices and cultures of federal science agencies. The autonomy of program managers to set priorities across fields or modify the decisions of external review groups can vary consider-

ably across agencies. The experiences of members of this committee when serving as members of advisory councils, advisory committees, and review panels are remarkably consistent, and also consistent with the views of program managers we have interviewed in this regard, although we have been unable to identify any systematic study on this topic.

In part, these apparent differences reflect the different histories and purposes of these agencies. DARPA, for example, was formed in the 1950s purposively as a small and flexible organization oriented toward revolutionary technology breakthroughs (Bonvillian and Sharp, 2001). Its use of advisory panels and review panels is flexible and ad hoc. By way of contrast, formalized peer review systems are core features of NSF and NIH, with a key distinction being that NSF program managers oversee both program development and the panel review process, whereas NIH separates responsibility for program development and operations from the review process. The peer review process at NIH operates primarily out of the Center for Scientific Review, which organizes review groups that often cut across programs and even institutes, and which generates ratings that are intended to evaluate proposals on a unitary scale that is the same across programs and institutes. This procedure makes it particularly difficult for a science manager at NIH to argue for overturning the results of the peer review process. As noted in Chapter 2, however, NIH science managers have greater discretion with funding instruments other than unsolicited proposals, which allow them to identify topics of interest and sometimes to set aside funds and create separate review processes for the solicited research.¹⁶

Proposals for quantifying of the benefits of research, as well as proposals for increased discretion for science managers, should be understood in the context of these conditions of influence and power. Quantification is sometimes presumed to reduce the influence of extramural scientists. If it has this effect, however, it does not necessarily increase the influence of the science managers who are closest to the scientific research programs. That effect will depend on how quantification is implemented and where in an agency the responsibility is placed for quantifying and for interpreting the results. We return to these issues in Chapters 5 and 6, where we discuss in more detail the use of quantitative analytic methods and deliberative processes for informing research assessment and priority setting.

NOTES

1. NSF's Division of Policy Research and Analysis became embroiled in a losing battle over the independence of its research findings, while ceasing to be a source of external research support (Greenberg, 2001); the congressional Office of Technology Assessment closed its doors in 1995.

2. This situation may be changing, as evidenced by recent statements by the president's science adviser (Marburger, 2005) and the NSF request to include a "new research effort to address policy-relevant science metrics" in its FY 2007 budget.
3. For example, Senator William Frist was quoted as saying in 2000, "[A]n improved process is needed for establishing goals and research priorities based on scientific data and health analysis, including moving beyond input measures and anecdotal evidence to develop new metrics to measure scientific advances and their causal relationship to improved health outcomes. Such measures will never be precise and should not be used as an absolute guide to determine where and how much to invest. Translating research along the continuum of basic, clinical, and applied research and, ultimately, to patient care almost always involves long periods; the linkages between these stages are seldom straightforward. Still, more comprehensive and transparent measurement tools would provide policy makers, the public, the scientific community, and patients with a more complete understanding of the role of government-sponsored research and help inform federal policy" (quoted in *Journal of the American Medical Association*, 2002). The comments of John Marburger, the science adviser to President George W. Bush, have already been noted.
4. According to one recent statement of this view from an NIH official: "Competitive pressures have pushed researchers to submit more conservative applications, and we must find ways to encourage greater risk-taking and innovation and to ensure that our study sections are more receptive to innovative applications" (Scarpa, 2006).
5. This report also proposes a mechanism to be used to assess the U.S. position. "Every five years, panels are convened to evaluate the fields in each major areas of science and technology (e.g., physics, biology, electrical engineering), their standing in the world, and the resources needed to reach and maintain world-class position. Evaluation focuses on outputs, such as important discoveries, and also on certain benchmarks of best practices, such as number of scientists and engineers and their training, or the current state of the laboratories and research facilities" (National Research Council, 1995a:15).
6. "This balance takes into consideration both the values of accountability associated with representative government and those of autonomy associated with an independent professional community. Not only does the government 'invest' in a public good . . . , but it delegates to other institutions the actual conduct of the research. It is thus the scientific community, as established in universities and other research institutions, that has responsibility for 'producing' research, discoveries, and new technologies" (Guston and Keniston, 1994:8).
7. Despite their seeming similarities, GPRA and PART reflect the different perspectives, budget priorities, and decision-making processes of the legislative and executive branches. OMB's initiation of PART, for example, has at times been described by OMB officials as representing dissatisfaction with the actual impact of GPRA on agency budgets. For its part, OMB's PART recommendations must still run the gauntlet of congressional appropriations committees. PART's impact on budget allocations remains problematical to date (Olsen and Levy, 2004).
8. In benefit-cost terms, their estimates indicate that an increase of \$100 billion for cancer research spent over 10 years would be "worthwhile if it had only a one-in-five chance of producing a 1 percent reduction in cancer mortality, and a four-in-five chance of producing nothing" (Murphy and Topel, 1999:3).
9. The rare isotope accelerator being proposed by some physicists for funding by the U.S. Department of Energy is estimated to cost \$1 billion. The National Aeronautics and Space Administration's construction of the James Webb Space Telescope has an estimated cost of \$4.5 billion.

10. Weingart (2005:120) observed, in discussing the difficulties of developing “correct” sets of bibliometric data, that the methodological and operational origins of the data can be “concealed from the end user who is not able to reflect upon the theoretical assumptions implied in their construction. . . . The healthy skepticism of years ago, albeit often for the wrong reasons, appears to have given way to an uncritical embrace of bibliometric measures and to an irresponsible use.”
11. A well-known instance of serendipity was the discovery that Viagra, initially developed to treat cardiac problems, had unanticipated effects on sexual performance.
12. Even for assessing “inventions” rather than “science discoveries,” patents have important limitations as performance measures. As noted by Jaffe and Trajtenberg (2002:3-4), “There are of course, important limitations to the use of patent data, the most glaring being the fact that not all inventions are patented. First, not all inventions meet the patentability criteria set by the USPTO, the United States Patent and Trademark Office (the invention has to be novel and nontrivial, and has to have commercial application). Second, the inventor has to make a strategic decision to patent as opposed to rely on secrecy or other means of appropriability.”
13. Roessner (2000:125) sees a choice as being imposed: “Posed as a choice, the question of quantitative versus qualitative methods or measures is a false one, at least to the professional evaluator residing in lofty isolation from the messy real world. . . . Legislators and other authoritative oversight bodies are increasingly asking public agencies for quantitative measures of research performance, and in so doing can generate all kinds of mischief.”
14. Erich Jarvis of Duke University, the 2002 recipient of the NSF’s Waterman Award, was quoted in *Science* (Mervis, 2004b:220) as saying “You learn the hard way not to send high-risk proposals to NSF or NIH, because they will get dinged by reviewers. Instead, you’re encouraged to tone down your proposal and request money for something you’re certain to be able to do.”
15. Researchers often seem to follow the dictum, “Thou shall not speak ill of a fellow [insert field]!” Consider, for example, the diplomatic response of Janez Potocnik, the European Union’s new commissioner for science and research on his appointment to the post: “when asked if any particular area of science has caught his interest since taking on the research job [Potocnik said that] (I)n practically all the areas you touch, you see interesting things going on” (Vogel, 2004).
16. Review processes also vary with funding instruments at NIH. The Center for Scientific Review is responsible for about two-thirds of peer review at NIH. Institute-based review groups do the bulk of the rest, such as reviewing proposals submitted in response to requests for applications.

4

Progress in Science

This chapter examines theories and empirical findings on the overlapping topics of progress in science and the factors that contribute to scientific discoveries. It also considers the implications of these findings for behavioral and social science research on aging. The chapter first draws on contributions from the history and sociology of science to consider the nature of scientific progress and the paths that lead to realizing the potential scientific and societal outcomes of scientific activity. It considers indicators that might be used to assess progress toward these outcomes. The chapter then examines factors that contribute to scientific discovery, drawing eclectically on the history and sociology of science as well as on theories and findings from organizational behavior, policy analysis, and economics.

THEORIES OF SCIENTIFIC PROGRESS

The history and sociology of science have produced extensive bodies of scholarship on some of these themes, generating in the process significant ongoing disagreements among scholars (see, e.g., Krige, 1980; Cole, 1992; Rule, 1997; Bowler and Morus, 2005). Most of this work focuses on processes and historical events in the physical and life sciences; relatively little of it addresses the social and behavioral sciences (or engineering, for that matter), except possibly subfields of psychology (e.g., Stigler, 1999). It is legitimate to ask whether this research even applies to the behavioral and social sciences (Smelser, 2005).¹

We do not attempt an encyclopedic coverage nor a resolution of the

debates, past and continuing, on such questions. Rather, we draw on this research to make more explicit the main issues underlying the tasks of prospective assessment of scientific fields for the purpose of setting priorities in federal research agencies, given the uncertain outcomes of research.

The history of science has produced several general theories about how science develops and evolves over long periods of time. A 19th century view is that of Auguste Comte, who argued that there is a hierarchy of the sciences, from the most general (astronomy), followed historically and in other ways by physics, chemistry, biology, and sociology. Sciences atop the hierarchy are characterized as having more highly developed theories; greater use of mathematical language to express ideas; higher levels of consensus on theory, methods, and the significance of problems and contributions to the field; more use of use theory to make verifiable predictions; faster obsolescence of research, to which citations drop off rapidly over time; and relatively fast progress. Sciences at the bottom of the hierarchy are said to exhibit the opposite characteristics (Cole, 1983).

Many adherents to this hierarchical view place the natural sciences toward the top of the hierarchy and the social sciences toward the bottom.² In this view, advances in the “higher” sciences, conceived in terms of findings, concepts, methodologies, or technologies that are thought to be fundamental, are held to flow down to the “lower” sciences, while the reverse flow rarely occurs. Although evidence of such a unidirectional flow from donor to borrower disciplines does exist (Losee, 1995), there are counterexamples. Historians and sociologists of science have offered evidence against several of these propositions, and particularly dispute the claimed association of natural science with the top of the hierarchy and social science with the bottom (e.g., Bourdieu, 1988; Cetina, 1999; Steinmetz, 2005). The picture is more complex, as noted below.

By far the best known modern theory of scientific progress is that of Thomas Kuhn (1962), which focuses on the major innovations that have punctuated the history of science in the past 350 years, associated with such investigators as Copernicus, Galileo, Lavoisier, Darwin, and Einstein. Science, in Kuhn’s view, is usually a problem-solving activity within clear and accepted frameworks of theory and practice, or “paradigms.” Revolutions occur when disparities or anomalies arise between theoretical expectation and research findings that can be resolved only by changing fundamental rules of practice. These changes occur suddenly, Kuhn claims, in a process akin to Gestalt shifts: in a relative instant, the perceived relationships among the parts of a picture shift, and the whole takes on a new meaning. Canonical examples include the Copernican idea that the Earth revolves around the Sun, Darwin’s evolutionary theory, relativity in physics, and the helical model of DNA.

A quite different account is that of John Desmond Bernal (1939). Inspired by Marxist social science and ideals of planned social progress, Bernal saw basic science progressing most vigorously when it was harnessed to practical efforts to serve humanity's social and economic needs (material well-being, public health, social justice). Whereas in Kuhn's view science progressed according to its inner logic, Bernal asserted that intellectual and practical advances could be engineered and managed.

Another tradition of thought, stemming from Derek Price's (1963) vision of a quantitative "science of science," has focused less on how innovations arise than on how they spread and how their full potential is exploited by small armies of scientists. Mainly pursued by sociologists of science, this line of analysis has focused on the social structure of research communities (e.g., Hagstrom, 1965), competition and cooperation in institutional systems (Merton, 1965; Ben-David, 1971), and structured communication in schools of research or "invisible colleges" (e.g., Crane, 1972). These efforts, while focused mainly on how science works, may imply principles for stimulating scientific progress and innovation.

There are also evolutionary models of scientific development, such as that of the philosopher David Hull (1988). Extending Darwin's account of evolution by variation and selection, Hull argues that scientific concepts evolve in the same way, by social or communal selection of the diverse work of individual scientists. In evolutionary views, science continually produces new ideas, which, like genetic mutations, are essentially unpredictable. Their ability to survive and expand their niches depends on environmental factors.

Bruno Latour and Steve Woolgar (1979) also offer an account of a selective struggle for viability among scientific producers. The vast majority of scientific papers quickly disappear into the maw of the scientific literature. The few that are used by other scientists in their work are the ones that determine the general direction of science progress. In evolutionary and competitive models, a possible function of science managers is to shape the environment that selects for ideas so as to propagate research that is judged to promote the agency's scientific and societal goals.

Stephen Cole (1992) emphasized a distinction between the frontier and the core of science that seems consistent with an evolutionary view. Work at the frontiers of sciences is characterized by considerable disagreement; as science progresses over time, disagreement decreases as processes such as empirical confirmation and paradigm shift select out certain ideas, while others become part of the received wisdom.

Although the view that different sciences have similar features at their respective frontiers is not unchallenged (Hicks, 2004), we have found the idea of frontier and core science to be useful in examining the extent to

which insights from the history and sociology of science, fields that have concentrated their attention predominantly on the natural sciences, also apply to the social and behavioral sciences.

Cole (1983, 1992) reports considerable evidence to suggest that different fields of science have similar features at the frontier, even if they are very different at the core. In the review of research proposals and journal submissions, an activity at the frontier of knowledge, he concludes that consensus about the quality of research is not systematically higher in the natural sciences than in the social sciences, citing the standard deviations of reviewers' ratings of proposals to the National Science Foundation, which were twice as large in meteorology as in economics.

In the core, represented by undergraduate textbooks, the situation appears to be quite different. Cole (1983) found that in textbooks published in the 1970s, the median publication date of the references cited in both physics and chemistry was before 1900, while the median publication date in sociology was post-1960. Sociology texts cited an average of about 800 references, while chemistry and physics texts generally cited only about 100. Moreover, a comparison of texts from the 1950s and the 1970s indicated that the material covered, as well as the sources cited, were much the same in both periods in physics and chemistry, whereas in sociology, the newer texts cited only a small proportion of the sources cited in the earlier texts.

Cole interpreted these findings as indicating that core knowledge in physics and chemistry was both more consensual and more stable over time than core knowledge in sociology. Such findings suggest that even though sciences may differ greatly at the core, for the purpose of assessing the progress of science at the frontiers of research fields, insights from the study of the natural sciences are likely to apply to the social sciences as well. They also point to the need to differentiate between “vitality,” as indicated by ferment at the frontier, and scientific progress as indicated by movement of knowledge from the frontier to the core.³ These findings suggest that the policy challenges for research managers making prospective judgments at the frontiers of research fields are quite similar across the sciences.

NATURE OF SCIENTIFIC PROGRESS

Scientific progress can be of various types—discoveries of phenomena, theoretical explanations or syntheses, tests of theories or hypotheses, acceptance or rejection of hypotheses or theories by the relevant scientific communities, development of new measurement or analytic techniques, application of general theory to specific theoretical or practical problems, development of technologies or useful interventions to improve human health and well-being from scientific efforts, and so forth. Consequently,

many different developments might be taken as indicators, or measures, of progress in science.

Science policy decision makers need to consider the progress and potential of scientific fields in multiple dimensions, accepting that the absence of detectable advance on a particular dimension is not necessarily evidence of failure or poor performance. Drawing on Weinberg's (1963) classification of internal and external criteria for formulating scientific choices, we make the practical distinction between internally defined types of scientific progress, that is, elements of progress defined by intellectual criteria, and externally defined types of progress, defined in terms of the contributions of science to society. Managers of public investments in science need to be concerned with both.

Scientific Progress Internally Defined

The literatures in the history of science and in science studies include various analyses and typologies of scientific and theoretical progress (e.g., Rule, 1997; Camic and Gross, 1998; Lamont, 2004). This section presents a distillation of insights from this research into a short checklist of major types of scientific progress. The list is intended as a reminder to participants in science policy decisions that assess the progress of scientific fields of the variety of kinds of progress science can make. Recognizing that these broad categories overlap and also that they are interdependent, with each kind of progress having the potential to influence the others, directly or indirectly, the list is intended to simplify a very complex phenomenon to a manageable level.

Types of Scientific Progress

Discovery. *Science makes progress when it demonstrates the existence of previously unknown phenomena or relationships among phenomena, or when it discovers that widely shared understandings of phenomena are wrong or incomplete.*

Analysis. *Science makes progress when it develops concepts, typologies, frameworks of understanding, methods, techniques, or data that make it possible to uncover phenomena or test explanations of them.* Thus, knowing where and how to look for discoveries and explanations is an important type of scientific progress. Improved theory, rigorous and replicable methods, measurement techniques, and databases all contribute to analysis.

Explanation. *Science makes progress when it discovers regularities in the ways phenomena change over time or finds evidence that supports, rules out, or leads to qualifications of possible explanations of these regularities.*

Integration. *Science makes progress when it links theories or explanations across different domains or levels of organization.* Thus, science progresses when it produces and provides support for theories and explanations that cover broader classes of phenomena or that link understandings emerging from different fields of research or levels of analysis.

Development. *Science makes progress when it stimulates additional research in a field or discipline, including research critical of past conclusions, and when it stimulates research outside the original field, including interdisciplinary research and research on previously underresearched questions. It also develops when it attracts new people to work on an important research problem.*

Recent scientific activities supported by the Behavioral and Social Research (BSR) Program of the National Institute on Aging (NIA) have yielded progress in the form of scientific advances of most of the above types. We cite only a few examples.

- *Discovery: The improving health of elderly populations.* An example is analyses of data from Sweden, which has the longest running national data set on longevity, that have shown that the maximum human life span has been increasing since the 1860s, that the rate of increase has accelerated since 1969, and that most of the change is due to improved probabilities of survival of individuals past age 70 (Wilmoth et al., 2000). Parallel trends have been discovered among the elderly in the form of declining physical disability, which declined in the United States from 26 percent of the elderly population in 1982 to 20 percent in 1999 (e.g., Manton and Gu, 2001), and declining cognitive impairment (e.g., Freedman et al., 2001, 2002). Such findings together suggest overall improvements in the health of elderly populations in high-income countries.

- *Analysis: Longitudinal datasets for understanding processes of aging.* The Health and Retirement Study (Juster and Suzman, 1995), a major ongoing longitudinal study that assesses the health and socioeconomic condition of aging Americans in which BSR played a central entrepreneurial role, has provided data that made possible, among other things, some of the discoveries about declining disability already noted. International comparative data sets on health risk factors and health outcomes, such as the Global Burden of Disease dataset (Ezzati et al., 2002), have also made significant scientific progress possible.

- *Explanation: Questioning and refining understandings.* Several BSR-funded research programs have yielded findings that called into question widely held views about aging processes. Examples include findings that question the beliefs that more health care spending leads to better health outcomes (Fisher et al., 2003a, 2003b), that increasing life expectancy im-

plies increased health care expenditures (Lubitz et al., 2003), that unequal access to health care is the main explanation for higher mortality rates among older people of lower socioeconomic status (e.g., Adda et al., 2003; Adams et al., 2003), and that aging is a purely biological process unaffected by personal or cultural beliefs (Levy, 2003). Other BSR-sponsored research has provided evidence that a previously noted association of depression with heart disease may be explained in part by a process in which negative affect suppresses immune responses (Rosenkranz et al., 2003).

- *Integration and development: Creating a biodemography of aging.* BSR supported and brought together “demographers, evolutionary theorists, genetic epidemiologists, anthropologists, and biologists from many different scientific taxa” (National Research Council, 1997:v) to seek coherent understandings of human longevity that are consistent with knowledge at levels from genes to populations and data from human and nonhuman species). This effort has helped to attract researchers from other fields into longevity studies, add vigor to this research field, and put the field on a broader and firmer interdisciplinary base of knowledge.

Paths to Scientific Progress

Scientific progress is widely recognized as nonlinear. Some new ideas have led to rapid revolutions, while other productive ideas have had lengthy gestation periods or met protracted resistance. Still other new ideas have achieved overly rapid, faddish acceptance followed by quick dismissal. An earlier generation of research in the history and sociology of science documented variety and surprise as characteristics of scientific progress, but it was not followed by broad transdisciplinary studies that developed and tested general theories of scientific progress.

No theory of scientific progress exists, or is on the horizon, that allows prediction of the future development of new scientific ideas or specifies how the different types of scientific progress influence each other—although they clearly are interdependent. Rather, recent studies by historians of science and practicing scientists typically emphasize the uncertainty surrounding which of a series of findings emerging at any point in time will be determinative of the most productive path for future scientific inquiries and indeed of the ways in which these findings will be used. Only in hindsight does the development of various experimental claims and theoretical generalizations appear to have the coherence that creates a sense of a linear, inexorable path.

Science policy seems to be in particular need of improved basic understanding of the apparently uncertain paths of scientific progress as a basis for making wiser, more efficient investments. Without this improved understanding, extensive investments into collecting and analyzing data on

scientific outputs are unlikely to provide valid predictors of some of the most important kinds of scientific progress. Political and bureaucratic pressures to plan for steady progress and to assess it with reliable and valid performance indicators will not eliminate the gaps in basic knowledge that must be filled in order to develop such indicators.

Despite the incompleteness of knowledge, the findings of earlier research remain a suggestive and potentially useful resource for practical research managers. They suggest a variety of state-of-knowledge propositions that are consistent with our collective experience on multiple advisory and review panels across several federal science agencies. We consider the following propositions worthy of consideration in discussions of how science managers can best promote scientific progress:

- Scientific discoveries are initially the achievements of individuals or small groups and arise in varied and largely unpredictable ways: the larger and more important the discoveries, the less predictable they would have been.
- The great majority of scientific products have limited impact on their fields; there are only a few major or seminal outputs. Whether or not new scientific ideas or methods become productive research traditions depends on an uncertain process that may extend over considerable time. Sometimes the impacts of research are quite different from those anticipated by the initial research sponsors, the researchers, or the individuals or organizations that first make use of it. For example, the Internet, which was developed as a means of fostering scientific communication among geographically dispersed researchers, has now become a leading channel for entertainment and retail business, among other things.
- Existing procedures for allocating federal research funds are most effective at the mid-level of scientific innovation, where there is consensus among established fields about the importance of questions and the direction and content of emerging questions in those fields.
- The uncertainties of scientific discovery and the difficulties of accurately identifying turning points and sharp departures in scientific inquiry suggest that research managers will do best with a varied portfolio of projects, including both mainstream and discontinuous or exploratory research projects. These uncertainties also suggest that assessment of a program's investments in research is most appropriately made at the portfolio rather than the project level.
- The portfolio concept also applies to a program's investments in analysis: in advancing the state of theoretical understanding, tools, and databases. Scientific progress in both the natural and social sciences may either follow or precede the development of new tools (instruments, models, algorithms, databases) that apply to many problems. Contrary to simple

models of scientific progress that have theory building as the grounding for empirical research or data collection as the foundation for theory building, the process is not linear or unidirectional.⁴ Program investments in theory building, tool development, and data collection can all contribute to scientific progress, but it is very difficult to predict which kinds of investments will be most productive at any given time (see National Research Council, 1986, 1988; Smelser, 1986).

- Scientific progress sometimes arises from efforts to solve technological or social problems in environments that combine concerns with basic research and with application. It can also arise in environments insulated from practical concerns. And progress can involve first one kind of setting and then the other (see Stokes, 1997).

Interdisciplinarity and Scientific Progress

The claim that the frontiers of science are generally located at the interstices between and intersections among disciplines deserves explicit attention because it is increasingly found in the conclusions and recommendations of national commissions and NRC committees (e.g., National Research Council, 2000b; Committee on Science, Engineering, and Public Policy, 2004) and in statements by national science leaders.⁵ Scholarship in the history and sociology of science is consistent with competing views on this claim. A considerable body of recent scholarship has noted that exciting developments often come at the edges of established research fields and at the boundaries between fields (Dogan and Pahre, 1990; Galison, 1999; Boix-Mansilla and Gardner, 2003; National Research Council, 2005b). Moreover, interdisciplinary thinking has become more integral to many areas of research because of the need to understand “the inherent complexity of nature and society” and “to solve societal problems” (National Research Council, 2005b:2).

The idea is that scientific advances are most likely to arise, or are most easily promoted, when scientists from different disciplines are brought together and encouraged to free themselves from disciplinary constraints. A good example to support this idea is the rapid expansion and provocative results of research on the biodemography of aging that followed the 1996 NRC workshop on this topic (National Research Council, 1997). The workshop occasioned serious efforts to develop and integrate related research fields.

To the extent that interdisciplinarity is important to scientific progress and for gaining the potential societal benefits of science, it is important for research managers to create favorable conditions for interdisciplinary contact and collaboration. In fact, for some time BSR has been seeking explicitly to promote both multidisciplinary and interdisciplinarity (Suzman, 2004).

For example, when the Health and Retirement Study was started in 1990, it was explicitly designed to be useful to economists, demographers, epidemiologists, and psychologists, and explicit efforts were made to convince those research communities that the study was not for economists only. BSR has reorganized itself and redefined its areas of interest on issue-oriented, interdisciplinary lines; sought out leading researchers and funded them to do what was expected to be ground-breaking and highly visible research in interdisciplinary fields; supported workshops and studies to define new interdisciplinary fields (e.g., National Research Council, 1997, 2000a, 2001c); created broadly based multidisciplinary panels to review proposals in emerging interdisciplinary areas; and funded databases designed to be useful to researchers in multiple disciplines for addressing the same problems, thus creating pressure for communication across disciplines. Some of the results, such as those already mentioned, have been notably productive and potentially useful.

The available studies seem to support the following conclusions about the favorable conditions for interdisciplinary science (Klein, 1996; Rhoten, 2003; National Research Council, 2005b):

- Successful interdisciplinary research requires both disciplinary depth and breadth of interests, visions, and skills, integrated within research groups.
- The success of interdisciplinary research groups depends on institutional commitment and research leadership with clear vision and team-building skills.
- Interdisciplinary research requires communication among people from different backgrounds. This may take extra time and require special efforts by researchers to learn the languages of other fields and by team leaders to make sure that all participants both contribute and benefit.
- New modes of organization, new methods of recruitment, and modified reward structures may be necessary in universities and other research organizations to facilitate interdisciplinary interactions.
- Both problem-oriented organization of research organizations and the ability to reorganize as problems change facilitate interdisciplinary research.
- Funding organizations may need to design their proposal and review criteria to encourage interdisciplinary activities.

Several conditions favorable to interdisciplinary collaboration can be affected by the actions of funders of research. For example, science agencies can encourage or require interdisciplinary collaboration in the research they support, support activities that specifically bring researchers together from different disciplines to address a problem of common interest, provide

additional funds or time to allow for the development of effective interdisciplinary communication in research groups or communities, and organize their programs internally and externally around interdisciplinary themes. They can ask review panels to consider how well groups and organizations that propose interdisciplinary research provide conditions, such as those above, that are commonly associated with successful interdisciplinary research. And they might also ensure that groups reviewing interdisciplinary proposals include individuals who have successfully led or participated in interdisciplinary projects.

Encouraging interdisciplinary research may have pitfalls, though. It is possible for funds to be offered but for researchers to fail to propose the kinds of interdisciplinary projects that were hoped for. Sometimes interdisciplinary efforts take hold, but they fail to produce important scientific advances or societal benefits. Interdisciplinarity can also become a mantra. If disciplines are at times presented as silos—*independent units with no connections among them*—interdisciplinary fields may also become silos that happen to straddle two fields. At any point in time, an observer can identify numerous new research trajectories, several involving novel combinations of existing disciplines. Thus, alongside recently institutionalized fields, such as biotechnology, materials science, information sciences, and cognitive (neuro)sciences, are claimants for scientific attention and programmatic support, such as vulnerability sciences, prevention science, and neuroeconomics.

Little is known about how to predict whether a new interdisciplinary field will take off in a productive way. Floral metaphors about budding fields are not always carried to the desired conclusion: many budding fields lack the intellectual or methodological germplasm to do more than pop up and quickly wither. It is at least as difficult to assess the prospects of interdisciplinary fields as of disciplinary ones, and probably more so (Boix-Mansilla and Gardner, 2003; National Research Council, 2005b).⁶

Federal agency science managers can act as entrepreneurs of interdisciplinary fields, so that their expansion from an interest of a small number of researchers into a recognizable cluster of activity may reflect the level of external support from federal agencies and foundations. As a field develops, though, a good indicator of vitality may be the exchange of ideas with other fields and particularly the export of ideas from the new field to other scientific fields or to practical use. But progress in interdisciplinary fields may be hard to determine from recourse to such indicators alone. Fields can be vital without exporting ideas to other fields. Policy analysis, now a well-established academic field of instruction and research, engages researchers from several social science disciplines, but it is a net importer of ideas (MacRae and Feller, 1998; Reuter and Smith-Ready, 2002).

It is worth noting that support for interdisciplinary research, although

it has unique benefits, may be a relatively high-risk proposition because it requires high-level leadership skills and innovative organizational structures. These characteristics of interdisciplinary research may pose special challenges for research managers in times of tightening budgets, when pressures for risk aversion may conflict with the need to develop innovative approaches to scientific questions and societal needs.

Contributions of Science to Society

In government agencies with practical missions, investments in science are appropriately judged both on internal scientific grounds and on the basis of their contributions to societal objectives. In the case of NIA, these objectives largely concern the improved longevity, health, and well-being of older people (National Institute on Aging, 2001). There are many ways research can contribute to these objectives. For simplicity, we group the societal objectives of science into four broad categories.

Identifying issues. *Science can contribute to society by identifying problems relating to the health and well-being of older people that require societal action or sometimes showing that a problem is less serious than previously believed.*

Finding solutions. *Science can contribute to society by developing ways to address issues or solve problems, for example, by improving prevention or treatment of diseases, improving health care delivery systems, improving access to health care, or developing new products or services that contribute to the longevity, health, or quality of life for older people in America.*

Informing choices. *Science can contribute to society by providing accurate and compelling information to public officials, health care professionals, and the public and thus promoting better informed choices about life and health by older people and better informed policy decisions affecting them.*

Educating the society. *Science can contribute to society by producing fundamental knowledge and developing frameworks of understanding that are useful for people facing their own aging and the aging of family members, making decisions in the private sector, and participating as citizens in public policy decisions. Science can also contribute by educating the next generation of scientists.*

Research on science utilization, a field that was most vital in the 1970s and that has seen some revival recently, has examined the ways in which scientific results, particularly social science results, may be used, particularly in government decisions (for recent reviews, see Landry et al., 2003, and Romsdahl, 2005, for some classic treatments, see Caplan, 1976; Weiss,

1977, 1979; Lindblom and Cohen, 1979). In terms of the above typology, this research mainly examines the use or nonuse of research results for informing choices by public policy actors. It does not much address the use of results by ordinary citizens, medical practitioners, the mass media, or other users involved in identifying issues and finding solutions, other than policy solutions. The most general classification in this research tradition of the ways social science is used is for enlightenment (i.e., providing a broad conceptual base for decisions) and as instrumental input (e.g., providing specific policy-relevant data). In addition, researchers note that social science results may be used to provide justification or legitimization for decisions already reached or as a justification for postponing decisions (Weiss, 1979; Oh, 1996; Romsdahl, 2005).

Federal science program managers face the challenges of establishing causal linkages between past research program activities and societal impacts and of projecting societal impacts from current and planned research activities. The challenges are substantial. Even when findings from social and behavioral science research influence policies and practices in the public and private sectors and may therefore be presumed to contribute to human well-being, they are seldom determinative. Indicators exist or could be created for many societal impacts of research (Cozzens et al., 2002; Bozeman and Sarewitz, 2005). In addition, evidence that the results of research are used, for example, in government decisions, may be considered an interim indicator of ultimate societal benefit, presuming that the decisions promote people's well-being.

Limits exist, however, to the ability of a mission agency to translate findings from the research it funds into practice. For the research findings of the National Institutes of Health (NIH) in general and NIA-BSR in particular, contributions to societal or individual well-being require the complementary actions of myriad other actors and organizations in government and the private sector, including state and local governments, insurance companies, nursing homes, physicians' practices, and individuals. According to Balas and Boren (2000:66), "studies suggest that it takes an average of 17 years for research evidence to reach clinical practice." Similarly lengthy processes and circuitous connections link research findings to more enlightened or informed policy making (Lynn, 1978).

A scientific development also may contribute to society in the above ways even if working scientists do not judge it to be a significant contribution on scientific grounds. For example, surveys sponsored by BSR produce data, for example on declining rates of disability among older people, that may be very useful for health care planning without, by themselves, contributing anything more to science than a phenomenon to be explained. Thus, it is appropriate for assessments of research progress to consider separately the effects of research activity on scientific and societal criteria. Scientific

activities and outputs may contribute to either of these two kinds of desirable outcomes or to both.

Interpreting Scientific Progress

The extent to which particular scientific results constitute progress in knowledge or contribute to societal well-being is often contested. This is especially the case when scientific findings are uncertain or controversial and when they can be interpreted to support controversial policy choices. Many results in applied behavioral and social science have these characteristics. Disagreements arise over which research questions are important enough to deserve support (that is, over which issues constitute significant social problems), about whether or not a finding resolves a scientific dispute or has unambiguous policy implications, and about many other aspects of the significance of scientific outputs. The more controversial the underlying social issues, the further such disagreements are likely to penetrate into the details of scientific method. Interested parties may use their best rhetorical tools to “frame” science policy issues and may even attempt to exercise power by influencing legislative or administrative decision makers to support or curtail particular lines of research.

These aspects of the social context of science are relevant for the measurement and assessment of scientific progress and its societal impact. They underline the recognition that the meaning of assessments of scientific progress may not follow in any straightforward way from the evidence the assessments produce. Assessing science, no matter how rigorous the methods that may be used, is ultimately a matter of interpretation. The possibility of competing interpretations of evidence is ever-present when using science indicators or applying any other analytic method for measuring the progress and impact of science. In Chapter 5, we discuss a strategy for assessing science that recognizes this social context while also seeking an appropriate role for indicators and other analytic approaches.

INDICATORS OF SCIENTIFIC PROGRESS

Research managers understandably want early indicators of scientific progress to inform decisions that must be made before the above types of substantive progress can be definitively shown. Although scientific progress is sometimes demonstrable very quickly, recent histories of science, as noted above, tend to emphasize not only the length of time required for research findings to generate a new consensus but also the uncertainties at the time of discovery regarding what precisely constitutes the nature of the discovery. Time lag and impact may depend on various factors, including the type of research and publication and citation practices in the field. A longitudinal

research project can be expected to take longer to yield demonstrable progress than a more conceptual project.

Research Vitality and Scientific Progress

Expressions of scientific interest and intellectual excitement, sometimes referred to as the vitality of a research field, have been suggested as a useful source of early indicators of scientific progress as defined from an internal perspective. Such indications of the development of science are of particular interest to science managers because many of them might potentially be converted into numerical indicators. They include the following:

- Established scientists begin to work in a new field.
- Students are increasingly attracted to a field, as indicated by enrollments in new courses and programs in the field.
- Highly promising junior scientists choose to pursue new concepts, methods, or lines of inquiry.
- The rate of publications in a field increases.
- Citations to publications in the field increase both in number and range across other scientific fields.
- Publications in the new field appear in prominent journals.
- New journals or societies appear.
- Ideas from a field are adopted in other fields.
- Researchers from different preexisting fields collaborate to work on a common set of problems.

Research on the nanoscale is an area that illustrates vitality by such indicators and that is beginning to have an impact on society and the economy. Zucker and Darby (2005:9) point to the rate of increase in publishing and patenting in nanotechnology since 1986 as being of approximately the same order of magnitude as the “remarkable increase in publishing and patenting that occurred during the first twenty years of the biotechnology revolution. . . . Since 1990 the growth in nano S&T articles has been remarkable, and now exceeds 2.5 percent of all science and engineering articles.” Major scientific advances are often marked by flurries of research activity, and many observers expect that such indications of research vitality presage major progress in science and applications.

However, research vitality does not necessarily imply future scientific progress. For example, research on cold fusion was vital for a time precisely because most scientists believed it would not lead to progress. In the social sciences, many fields have shown great vitality for a period of time, as indicated by numbers of research papers and citations to the central works, only to decline rapidly in subsequent periods. Rule (1997), in his study

of progress in social science, discusses several examples from sociology, including the grand social theory of Talcott Parsons (1937, 1964), ethnomethodology (e.g., Garfinkel, 1967), and interaction process analysis (e.g., Bales, 1950). Although these fields were vital for a time, in longer retrospect many observers considered them to have been far less important to scientific progress than they had earlier appeared to be. Rule suggests several possible interpretations of this kind of historical trajectory: the fields that looked vital were in fact intellectual dead-ends; research in the fields did make important contributions that were so thoroughly integrated into thinking in the field that they became common knowledge and were no longer commonly cited; and the fields represented short-term intellectual tastes that lost currency with a shift in theoretical concerns. With enough hindsight, it may be possible to decide which interpretation is most correct, although disagreements remain in many specific cases. But the resource allocation challenge for a research manager, given multiple alternative fields whose aggregate claims for support exceed his or her program budget, is to make the correct interpretation of research vitality prospectively: that is, to project whether the field will be judged in hindsight to have produced valuable contributions or to have been no more than a fad or an intellectual dead-end.

Another trajectory of research is problematic for research managers who would use vitality as an indicator of future potential. Some research findings or fields lie dormant for considerable periods without showing signs of vitality, before the seminal contributions gain recognition as major scientific advances. Such findings have been labeled as “premature discoveries” (Hook, 2002) and “sleeping beauties” (van Raan, 2004b). These are not findings that are resisted or rejected; rather, they are unappreciated, or their uses or implications are not initially recognized (Stent, 2002). In effect, the contribution of such discoveries to scientific progress or societal needs or both lies dormant until there is some combination of independent discoveries that reveal the potency of the initial discovery. In such cases, vitality indicators focused predominately on the discovery and its related line of research would have been misleading as predictors of long-term scientific importance.

An instructive example of the limitations of vitality measures as early indicators in the social sciences is the intellectual history of John Nash’s approach to game theory—an approach that was recognized, applied, and then dismissed as having limited utility, only to reemerge again as a major construct (the Nash equilibrium), not only in the social and behavioral sciences but also in the natural sciences. As recounted by Nasar (1998), the years following Nash’s seminal work at RAND in the early 1950s were a period of flagging interest in game theory. Luce and Raiffa’s authoritative overview of the field in 1957 observed: “We have the historical fact that many social scientists have become disillusioned with game theory. Initially

there was a naïve band-wagon feeling that game theory solved innumerable problems of sociology and economics, or that, at least it made their solution a practical matter of a few years' work. This has not turned out to be the case" (quoted in Nasar, 1998:122). In later retrospect, game theory became widely influential in the social and natural sciences, and Nash was awarded the Nobel Memorial Prize in Economics in 1994.

The complexity of the relationship between the quantity of scientific activity being undertaken during a specific period and the pace of scientific progress (or the rate at which significant discoveries are made) can perhaps be illustrated by analogy to a bicycle race: a group of researchers, analogous to the peloton or pack in a bicycle race, proceeds along together over an extended period until a single individual or a small group attempts a break-away to win the race. Some breakaways succeed and some fail, but because of the difficulties of making progress by working alone (wind resistance, in the bicycle race analogy), individuals need the cooperation of a group to make progress over the long run and to create the conditions for racing breakaways or scientific breakthroughs. When scientific progress follows this model, fairly intense activity is a necessary but not sufficient condition for progress. Alternatively, the pack may remain closely clustered together for extended periods of time, advancing apace yet with a sense that little progress toward victory, however specified, is being made (Horan, 1996).

In our judgment, these various trajectories of scientific progress imply that *quantitative indicators, such as citation counts, require interpretation if they are to be used as part of the prospective assessment of fields*. Moreover, the implications of intensified activity in a research area may be quite different depending on the mission goals and the perspective of the agency funding the work. Significant research investments can create activity in a field by encouraging research and supporting communication among communities of researchers. But activity need not imply progress, at least not in terms of some of the indicators listed above, such as the export of ideas to other fields. If research managers conflate the concepts of scientific activity and progress, they can create self-fulfilling prophecies by simply creating scientific activity. These warnings become increasingly important as technical advances in data retrieval and mining make it easier to create and access quantitative indicators of research vitality and as precepts of performance assessment increase pressures on research managers to use quantitative indicators to assess the progress and value of the research they support.

Indicators of Societal Impact

A variety of events may indicate that scientific activities have generated results that are likely to have practical value, even though such value may not (yet) have been realized. Such events might function as leading indica-

tors of the societal value of research. These events typically occur outside research communities. For example:

- Research is cited as the basis for patents that lead to licenses.
- Research is used to justify policies or laws or cited in court opinions.
- Research is prominently discussed in trade publications of groups that might apply it.
- Research is used as a basis for practice or training in medicine or other relevant fields of application.
- Research is cited and discussed in the popular press as having implications for personal decisions or for policy.
- Research attracts investments from other sources, such as philanthropic foundations.

Some of these potential indicators are readily quantifiable, so, like bibliometric indicators, they are attractive means by which science managers can document the value of their programs. But as with quantitative indicators of research vitality, the meaning of quantitative indicators of societal impact is subject to differing interpretations. For example, as studies of science utilization have emphasized, the use of research to justify policy changes may mean that the research has changed policy makers' thinking or only that it provides legitimation for previously determined positions. Moreover, policy makers have been known to use research to justify a policy when the relevant scientific community is in fact sharply divided about the importance or even the validity of the cited research. Such research nevertheless has societal impact, even if not of the type the scientists may have expected.

FACTORS THAT CONTRIBUTE TO SCIENTIFIC DISCOVERIES

Historically, analysis of the factors that contribute to scientific discoveries has occurred at least at three different levels of analysis. Macro-level studies have considered the effects of the structures of societies—their philosophical, social, political, religious, cultural, and economic systems (Hart, 1999; Jones, 1988; Shapin, 1996). Meso-level analyses have examined the effects of functional and structural features of “national research and innovation systems”—for example, the relative apportionment of responsibility and public funding for scientific inquiry among government entities, quasi-independent research institutes, and universities (Nelson, 1993). Micro-level studies have examined the associations between indicators of progress and such factors as the organization of research units and the age of the researcher (Deutsch et al., 1971).

The programmatic latitude of any single federal science unit to adjust

its actions to promote scientific discovery relates almost exclusively to micro-level factors. Even then, agency policies, legislation, and higher level executive branch policies may limit an agency's options. For this reason, we look most closely at micro-level factors. It is nevertheless worth examining the larger structural factors affecting conditions for scientific discovery, if only to understand the implicit assumptions likely to be accepted by BSR's advisers and staff.

A convenient means of documenting contemporary thinking on the factors that contribute to scientific advances is to examine the series of "benchmarking" studies of the international standing of U.S. science in the fields of materials science, mathematics, and immunology made by panels of scientists under the auspices of the National Academies' Committee on Science, Engineering, and Public Policy (COSEPUP). The benchmarking was conducted as a methodological experiment in response to a series of studies that had sought to establish national goals for U.S. science policy and to mesh these goals with the performance reporting requirements of the Government Performance and Results Act (Committee on Science, Engineering, and Public Policy, 1993, 1999a; National Research Council, 1995a).

The benchmarking reports covered the fields of mathematics (Committee on Science, Engineering, and Public Policy, 1997), materials science (Committee on Science, Engineering, and Public Policy, 1998), and immunology (Committee on Science, Engineering, and Public Policy, 1999b); they represented attempts to assess whether U.S. science was achieving the stated goals of the National Goals report (Committee on Science, Engineering, and Public Policy, 1993) that the United States should be among the world leaders in all major areas of science and should maintain clear leadership in some major areas of science. These reports can be used to infer the collective beliefs across a broad range of the U.S. scientific community about the factors that contribute to U.S. scientific leadership, and implicitly to the factors that foster major scientific discoveries. The reports are also of interest because several of the factors they cite—for example, initiation of proposals by individual investigators, reliance on peer-based merit review—are the cynosures of proposals to modify the U.S. science system.

Across the three benchmarking reports, the core repeatedly cited as necessary for scientific progress was adequate facilities, quality and quantity of graduate students attracted to a field (and their subsequent early career successes in the field), diversity in funding sources, and adequate funding. In addition, with regard to the comparative international strength and the leadership position of U.S. science in these fields, the reports placed special emphasis on the "structure and financial-support mechanisms of the major research institutions in the United States" and on its organization of higher education research (Committee on Science, Engineering and Public Policy, 1999b:35). Also highlighted as a contributing factor in "fostering

innovation, creativity and rapid development of new technologies” was the “National Institutes of Health (NIH) model of research-grant allocation and funding: almost all research (except small projects funded by contracts) is initiated by individual investigators, and the decision as to merit is made by a dual review system of detailed peer review by experts in each subfield of biomedical science” (p. 36).⁷

We accept the proposition that adequate funds to support research represents a necessary condition for sustained progress in a scientific field. Research progress also depends on the supply of researchers (including the number, age, and creativity of current and prospective researchers) and the organization of research, including the number and disciplinary mix of researchers engaged in a project or program and structure of the research team.

Supply of Researchers

The number, creativity, and age distribution of researchers in a field together affect the pace of scientific progress in the field. Numbers are important to the extent that the ability to generate scientific advances is randomly distributed through a population of comparably trained researchers. Fields with a larger number of active researchers can be expected to generate more scientific advance than fields with smaller such numbers. The pace of scientific advance across fields presumably also varies with their ability to attract the most able/creative/productive scientists. The attractiveness of a field at any point in time is likely to depend on its intellectual excitement (the challenges of the puzzles that it poses), its societal significance, the resources flowing to it to support research, and the prospects for longer term productive and gainful careers. Fields that exhibit these characteristics are likely to attract relatively larger cohorts of younger scientists; if scientific creativity is inversely correlated with age, such fields may be expected to exhibit greater vitality than those with aging cohorts of scientists.

This view is supported by much expert judgment and a number of empirical studies. For example, a study by the National Research Council (1998:1) noted that “The continued success of the life-science research enterprise depends on the uninterrupted entry into the field of well-trained, skilled, and motivated young people. For this critical flow to be guaranteed, young aspirants must see that there are exciting challenges in life science research and they need to believe that they have a reasonable likelihood of becoming practicing independent scientists after their long years of training to prepare for their careers.”

Career opportunities for scientists affect the flow of young researchers into fields. Recent studies of career opportunities in the life sciences have noted that a “crisis of expectations” arises when career prospects fall short

of scientific promise (Freeman et al., 2001). Similar observations have been made at other times for the situations in physics, mathematics, computer science, and some fields of engineering. Studies also point, in general, to a decline in research productivity around midcareer. As detailed by Stephan and Levin (1992), the decline reflects influences on both the willingness and ability of researchers to do scientific research. Older scientists are also seen to be slower to accept new ideas and techniques than are younger scientists.⁸

Organization of Research

Since World War II, the social contract by which the federal government supports basic research has involved channeling large amounts of this support through awards to universities, much of that through grants to individual investigators. It is appropriate to consider whether such choices continue to be optimal and to consider related questions concerning the determinants of the research performance of individual faculty and of specific institutions or sets of institutions (Guston and Keniston, 1994; Feller, 1996).

As detailed above, U.S. support of academic research across many fields, including aging research, is predicated on the proposition that “little science is the backbone of the scientific enterprise. . . . For those who believe that scientific discoveries are unpredictable, supporting many creative researchers who contribute to S&T, or the science base is prudent science policy” (U.S. Congress Office of Technology Assessment, 1991:146). Against this principle, trends toward “big science” and the requirements of interdisciplinary research have opened up the question of the optimal portfolio of funding mechanisms and award criteria to be employed by federal science agencies. Of special interest here as an alternative to the traditional model of single investigator–initiated research are what have been termed “industrial” models of research (Ziman, 1984) or Mode II research; that is, research undertakings characterized by collaboration or teamwork among members of research groups participating in formally structured centers or institutes. Requests for proposals directed toward specific scientific, technological, and societal objectives; initiatives supporting collaborative, interdisciplinary modes of inquiry organized as centers rather than as single principal investigator projects; and use of selection criteria in addition to scientific merit are by now well-established parts of the research programs of federal science agencies, including NIH and the National Science Foundation.⁹

A recurrent issue for federal science managers and for scientific communities is the relative rate of return to alternative arrangements, such as funding mechanisms. Making such comparisons is challenging. First, different research modes (e.g., single investigator–initiated proposals and multidisciplinary, center-based proposals submitted in response to a Request

for Application) may produce different kinds of outputs. Single-investigator awards, typically described as the backbone of science, are intended cumulatively to build a knowledge base that affects clinical practice or public policy, to support the training of graduate students, to promote the development of networks of researchers and practitioners, and more—but no single awardee is expected to do all these things. Center awards also are expected to contribute to scientific progress—indeed to yield “value added” above the progress that can come from multiple single-investigator awards—but unlike single-investigator awards, they are typically expected to devote explicit attention to the other outcomes, such as translating the results of basic research into clinical practice. Because different modes of research support are expected to support different mixes of program objectives, direct comparisons of “performance” or “productivity” between or among them involves a complex set of weightings and assessments, both in terms of defining and measuring scientific progress and in assigning weights to the different kinds of scientific, programmatic, and societal objectives against which research is evaluated.

Little empirical evidence exists to inform comparisons among modes of research support. Empirical studies, most frequently in the form of bibliometric analyses, exist to compare the productivity of interdisciplinary research units, but these studies are not designed to answer the question of how much scientific progress would have been achieved had the funds allocated to such units been apportioned instead among a larger and more diverse number of single investigator awards (Feller, 1992). Detailed criteria, for example, have been advanced to evaluate the performance of NIH’s center programs (Institute of Medicine, 2004), and a number of center programs have been evaluated. However, these evaluations have not added up to a systematic assessment.¹⁰

Expert judgment, historical assessment, and analysis of trends in science provide some support for core propositions about the sources of the vitality of U.S. science: adequate and sustainable funding; multiple, decentralized, funding streams; strong reliance on investigator-initiated proposals selected through competitive, merit-based review; coupling basic research with graduate education; and supplementary funding for capital-intensive modes of inquiry, interdisciplinary collaboration, targeted research objectives, and translation of basic research findings into clinical practice or technological innovations. Still, these principles may not provide wise guidance for the support of behavioral and social science research on aging, for three reasons. First, these observations come from experience with the life sciences, engineering sciences, and physical sciences, and it is not known whether the dynamics of scientific inquiry and progress are the same in the social and behavioral sciences. Second, it is not known whether recent trends in scientific inquiry, such as in the direction of interdisciplinarity, will continue,

stop, or soon lead to a fundamental transformation in the way in which cutting-edge science (including in research on aging) is done. Third and perhaps most important, applying these principles presumes an environment of increasing total funds for research. In the more austere budget environment now projected for NIH and its subunits, it will not be possible to increase funding for all modes of support. Turning to existing research for guidance may prove of limited value for making trade-offs among competing funding paradigms.

IMPLICATIONS FOR DECISION MAKING

1. *No theory exists that can reliably predict which research activities are most likely to lead to scientific advances or to societal benefit.* The gulf between the decision-making environment of the research manager and the historian or other researcher retrospectively examining the emergence and subsequent development of a line of research is reflected in Weinberg's (2001:196) observation, "In judging the nature of scientific progress, we have to look at mature scientific theories, not theories at the moments when they are coming into being." The history of science shows that evidence of past performance and current vitality, that is, of interest among scientists in a topic or line of research, are imperfect predictors of future progress. Thus, although it seems reasonable to expect that a newly developing field that generates excitement among scientists from other fields is a good bet to make progress in the near future, this expectation rests more on anecdote than on systematic empirical research. Notwithstanding the continuing search for improved quantitative measures and indicators for prospective assessment of scientific fields, practical choices about research investments will continue to depend on judgment. We address the prospects and potential roles of quantitative and other methods of science assessment in Chapter 5.

2. *Science produces diverse kinds of benefits; consequently, assessing the potential of lines of research is a challenging task.* Assessments should carefully apply multiple criteria of benefit. Science proceeds toward improving understanding and benefiting society on several fronts, but often at an uneven pace, so that a line of research may show rapid progress on one dimension or by one indicator while showing little or no progress on another. In setting research priorities among lines of research, it is important to consider evidence of past accomplishments on the several dimensions of scientific advances (discovery, analysis, explanation, integration, and development) and of contributions to society (e.g., identifying issues, finding solutions, informing choices).

The policy implications of a finding that a line of research is not currently making much progress on one or more dimensions are not self-evident.

Such an assessment might be used as a rationale for decreasing support (because the funds may be expected to be poorly spent), for increasing support (for example, if the poor performance is attributed to past underfunding), or for making investments to redirect the field so as to reinvigorate it. A field that appears unproductive may be stagnant, fallow, or pregnant. Telling which is not easy. Judgment can be aided by the assessments of people close to the field, although not just those so close as to have a vested interest in its survival or growth. The same kind of advice is useful for judging the proper timing for efforts to invest in fields in order to keep them alive or to reinvigorate them.

3. *Portfolio diversification strategies that involve investment in multiple fields and multiple kinds of research are appropriate for decision making, considering the inherent uncertainties of scientific progress.* Through such strategies, research managers can minimize the consequences of overreliance on any single indicator of research quality or progress or any single presumption about what kinds of research are likely to be most productive. It is appropriate to diversify along several dimensions, including disciplines, modes of support, emphasis on theoretical or applied objectives, and so forth. Diversification is also advisable in terms of the kinds of evidence relied on to make decisions about what to support. For example, when quantitative indicators and informed peer judgment suggest supporting different lines of research, it is worth considering supporting some of each.

4. *Research managers should seek to emphasize investing where their investments are most likely to add value. This consideration may affect emphasis on types of scientific progress, research organizations and modes of support, and areas of support.*

a. *Types of scientific progress.* Even as they continue to pursue support of major scientific and programmatic advances, research managers may also find it productive to support improvements in databases and analytic techniques, efforts to integrate knowledge across fields and levels of analysis, efforts to examine underresearched questions, and the entry of new people to work on research problems.

b. *Research organizations and modes of support.* Research managers should consider favoring support to research organizations or in modes that have been shown to have characteristics that are likely to promote progress, either generally or for specific fields or lines of scientific inquiry. NIH has multiple funding mechanisms available that would allow support for particular types of organizations (Institute of Medicine, 2004). An ongoing study by Hollingsworth (2003:8) identifies six organizational characteristics as “most important in facilitating the making of major discoveries” (see Box 4-1). Research managers might consider the findings of such studies in making choices about what kinds of organizations to support, especially in efforts to promote scientific innovation.

BOX 4-1

Characteristics of Organizations That Produced Major Biomedical Discoveries: The Hollingsworth Study

Rogers Hollingsworth and colleagues (Hollingsworth and Hollingsworth, 2000; Hollingsworth, 2003) have been examining the characteristics of biomedical research organizations associated with the production of major discoveries in the 20th century. The study relied on the scientific community to define major discoveries by designating as “major” discoveries that resulted in the awarding of a Nobel prize in chemistry or physiology or medicine, an Arthur and Mary Lasker prize, a Louisa Gross Horwitz prize, a Crafoord prize, or a Copley medal or that received 10 nominations for a Nobel prize in any three years before 1941. The study compared 20 research organizations in the United States in which two or more major discoveries occurred with 100 other research organizations. It identified organizational characteristics that facilitate or hamper the making of major discoveries, suggesting some directions that research organizations might take if they want to maintain or increase their likelihood of nurturing major discoveries. The characteristics are

- Organizational autonomy (the capacity to make decisions according to criteria the organization develops “independently of external disciplinary norms and governing authorities”).
- Organizational flexibility to shift research areas quickly.
- Moderate scientific diversity, with depth in various fields and in theory, methods, and instrumentation, as well as internalized scientific diversity among a large portion of the scientific staff.
- Frequent and intense interaction among scientists with different cognitive perspectives (e.g., joint publications, journal clubs, and sharing meals and leisure time activities).
- Leaders with a strategic vision for integrating diverse areas and focusing research, the ability to secure funding and recruit diverse personnel, the ability to “provide rigorous criticism in a nurturing environment,” and the capacity to orchestrate a diverse group of scientists while orienting them to future directions.
- Recruitment of “scientists who internalize moderately high levels of diversity at the time of their appointments.”

The study also suggests three organizational characteristics that hamper the making of major discoveries:

- Sharp differentiation of boundaries among scientific areas, with many departments and fundraising and recruitment delegated to them.
- Centralized, hierarchical authority and bureaucratic procedures.
- Hyperdiversity that precludes effective communication across fields.

c. *Areas of support.* Some fields may have sufficient other sources of funds that they do not need NIA support, or only need small investments from NIA to leverage funds from other sources. In other fields, however, BSR may be the only viable sponsor for the research. BSR managers may reasonably choose to emphasize supporting research in such fields because of the unlikelihood of leveraging funds. The value-added issue also affects decisions on modes of support and types of research to support.

5. *Interdisciplinary research.* BSR should continue to support issue-focused interdisciplinary research to promote scientific activities and collaborations related to its mission that might not emerge from existing scientific communities and organizations structured around disciplines. Interdisciplinary research has significant potential to advance scientific objectives that research management can promote, such as scientific integration and development and scientists' attention to societal objectives of science consistent with BSR's mission. Moreover, BSR has a good track record of promoting these objectives through its support of selected areas of interdisciplinary, issue-focused research.

BSR should continue to solicit research in areas that require interdisciplinary collaboration, to support data sets that can be used readily across disciplines, to fund interdisciplinary workshops and conferences, and to support cross-institution, issue-focused interdisciplinary research networks. Supporting such research requires special efforts and skills of research managers but holds the promise of yielding major advances that would not come from business-as-usual science.

NOTES

1. It is often argued that progress in the behavioral and social sciences is qualitatively different from progress in the natural sciences. As noted in a National Research Council review of progress in the behavioral and social sciences (Gerstein, 1986:17), "Because they are embedded in social and technological change, subject to the unpredictable incidence of scientific ingenuity and driven by the competition of differing theoretical ideas, the achievements of behavioral and social science research are not rigidly predictable as to when they will occur, how they will appear, or what they might lead to." The unstated (and untested) implication is that this unpredictability is more characteristic of the social sciences than the natural sciences. Another view states: "In the natural sciences, a sharp division of labor between the information-gathering and the theory-making functions is facilitated by an approximate consensus on the definition of research purposes and on the conceptual economizers guiding the systematic selection and organization of information. In the social sciences, where the subject matter of research and the comparatively lower level of theoretical agreement generally do not permit comparable consensus on the value and utility of information extracted from phenomena, sharp division of labor between empirical and theoretical tasks is less warranted" (Ezrahi, 1978:288). Even the same techniques are thought to have quite different roles in the social and natural sciences: "The role of statistics in social science is thus fundamentally different from its role in much of the physical science, in that it creates and defines the objects of study much

- more directly. Those objects are no less real than those of the physical science. They are even more often much better understood. But despite the unity of statistics—the same methods are useful in all areas—there are fundamental differences, and these have played a role in the historical development of all these fields” (Stigler, 1999:199).
2. Some observers even question the claims of the behavioral and social sciences to standing as sciences. As observed in a recent text on the history of science, “In the end, perhaps the most interesting question is: Did the drive to create a scientific approach to the study of human nature achieve its goal? For all the money and effort poured into creating a body of practical information on the topic, many scientists in better established areas remain suspicious, pointing to a lack of theoretical coherence that undermines the analogy with the ‘hard’ sciences” (Bowler and Morus, 2005:314-315).
 3. According to Cole (2001:37), “The problem with fields like sociology is that they have virtually no core knowledge. Sociology has a booming frontier but none of the activity at that frontier seems to enter the core.”
 4. As noted by Galison (1999:143), “Experimentalists . . . do not march in lockstep with theory. . . . Each subculture has its own rhythms of change, each has its own standards of demonstration, and each is embedded differently in the wider culture of institutions, practices, inventions and ideas.”
 5. Rita Colwell, former director of the National Science Foundation, has stated that “Interdisciplinary connections are absolutely fundamental. They are synapses in this new capability to look over and beyond the horizon. Interfaces of the sciences are where the excitement will be the most intense” (Colwell, 1998).
 6. As stated in a recent National Research Council (2005b:150) report, “A remaining challenge is to determine what additional measures, if any, are needed to assess interdisciplinary research and teaching beyond those shown to be effective in disciplinary activities. Successful outcomes of an interdisciplinary research (IDR) program differ in several ways from those of a disciplinary program. First, a successful IDR program will have an impact on multiple fields or disciplines and produce results that feed back into and enhance disciplinary research. It will also create researchers and students with an expanded research vocabulary and abilities in more than one discipline and with an enhanced understanding of the interconnectedness inherent in complex problems.”
 7. Consistent with the belief that competitive, merit-based review is key to creating the best possible conditions for scientific advance is the articulation of how “quality” is to be achieved and gauged under the Research and Development Investment Criteria established by the Office of Science and Technology Policy and the Office of Management and Budget on June 5, 2005: “A customary method for promoting quality is the use of a competitive, merit-based process” (<http://www.whitehouse.gov/omb/memoranda/m03-15.pdf>, p. 7).
 8. As Max Planck famously remarked, “a new scientific truth does not triumph by convincing its opponents and making them see the light, but because the its opponents eventually die, and a new generation grows up that is familiar with it.” Stephan and Levin (1992:83) write: “empirical studies of Planck’s principle for the most part confirm the hypothesis that older scientists are slower than their younger colleagues are to accept new ideas and that eminent older scientists are the most likely to resist. The operative factor in resistance, however, is not age per se but, rather, the various indices of professional experience and prestige correlated with age . . . [Y]oung scientists . . . may also be less likely to embrace new ideas, particularly if they assess such a course as being particularly risky.” Thus, a graying scientific community affects the rate of scientific innovation directly by being less productive and indirectly by being slow to accept new ideas as they emerge.
 9. Interdisciplinary research and the industrial model of research are often found together, but they are not identical. One may organize centers based primarily on researchers from a single discipline, and researchers from several disciplines may collaborate, as co-principal investigators or as loosely coupled teams, on one-time awards. At NIH, research center

grants “are awarded to extramural research institutions to provide support for long-term multidisciplinary programs of medical research. They also support the development of research resources, aim to integrate basic research with applied research and transfer activities, and promote research in areas of clinical applications with an emphasis on intervention, including prototype development and refinement of products, techniques, processes, methods, and practices” (Institute of Medicine, 2004).

10. “NIH does not have formal regular procedures or criteria for evaluating center programs. From time to time, institutes conduct internal program reviews or appoint external review panels, but these ad hoc assessments are usually done in response to a perception that the program is no longer effective or appropriate rather than as part of a regular evaluation process. Most of these reviews rely on the judgment of experts rather than systematically collected objective data, although some formal program evaluations have been performed by outside firms using such data” (Institute of Medicine, 2004:121).

5

Methods of Assessing Science

This chapter presents a general framework for thinking about methods of assessing science retrospectively or prospectively, reviews the conceptual and empirical literatures on the selected methods, and discusses their likely relevance and feasibility for research priority-setting decisions in the Behavioral and Social Research (BSR) Program of the National Institute on Aging (NIA). The focus here is on seeking methods that can provide science managers with the best possible input to priority-setting decisions while also achieving basic goals of accountability and rational decision making. Quantitative methods are attractive in terms of accountability, in the accountant's sense of comparing different investments in research on a common numerical scale. They are also conducive to improved outcomes to the extent that the measures are valid indicators of what they purport to measure.

Most of this chapter is devoted to examining the strengths and weaknesses of various methods of assessing science. Science assessments have become commonplace and include assessments of fundamental science (National Science and Technology Council, 1996), of technology development programs (Link, 1996; Ruegg and Feller, 2003), and of the performance of specific academic and research laboratories, in both the United States and other countries (e.g., Bozeman and Melkers, 1993; Moed et al., 2004). The assessments include commission reports, agency-specific commissioned evaluations, and academic works. As Table 5-1 shows, various assessment methodologies have arisen from different disciplinary and multidisciplinary perspectives, measuring different aspects of scientific activity, and addressing various science and technology policy and assessment questions.

TABLE 5-1 Some Methodologies for Science Assessment and the Attributes of Scientific Activity They Measure

Methods	Attributes Measured
Coauthorship links, multinational research articles	Scientific collaboration, globalization
Patent citation analysis	Economic value of patents
Cross-disciplinary coauthorships and citations	Multidisciplinarity and interdisciplinarity of research
Citations from clinical guidelines, regulations, and newspapers	Practical use of research
Scientist-inventor relationships, citations from articles to patents	Knowledge flows from science to technology
Co-occurring word and citation analysis	Sociocognitive structures in science
Use of first names of authors or inventors	Participation of women in science

SOURCE: Moed et al. (2004).

These assessment efforts have generated several broadly accepted “best practice” principles. For example, the National Science and Technology Council (1996:xii) set forth the following nine principles for assessment of fundamental science programs:

- Begin with a clearly defined statement of program goals.
- Develop criteria intended to sustain and advance the excellence and responsiveness of the research system.
 - Establish performance indicators that are useful to managers and encourage risk-taking.
 - Avoid assessments that would be inordinately burdensome or costly or that would create incentives that are counterproductive.
 - Incorporate merit review and peer evaluation of program performance.
 - Use multiple sources and types of evidence, for example, a mix of quantitative and qualitative indicators and narrative text.
 - Experiment in order to develop an effective set of assessment tools.
 - Produce assessment reports that will inform future policy development and subsequent refinement of program plans.
 - Communicate results to the public and elected representatives.

These principles are generally sensible, but they leave some important questions unaddressed. One of these is how to establish useful performance

indicators and incorporate peer review and evaluation at the same time. In this chapter, we adopt a conceptual framework for thinking about assessment methods that we think will allow such questions to be addressed more systematically. Our recommendations are in Chapter 6.

A FRAMEWORK: ANALYSIS AND DELIBERATION AS ASSESSMENT STRATEGIES

We find it useful to consider the issues of research assessment, both prospective and retrospective, in light of a distinction made in a previous National Research Council (NRC) study. In *Understanding Risk: Informing Decisions in a Democratic Society* (1996), an NRC committee distinguished between two methods for seeking practical understanding that it called analysis and deliberation. Analysis “uses rigorous, replicable methods developed by experts to arrive at answers to factual questions”; deliberation “uses processes such as discussion, reflection, and persuasion to communicate, raise and collectively consider issues, increase understanding, and arrive at substantive decisions” (p. 20). In stylized terms, counting patents or citations to studies, constructing network diagrams of communication patterns, and enumerating publications in designated major journals are analytic methods, whereas peer review conducted through discussions in advisory panels is a deliberative method.

Understanding Risk noted that science policy decisions typically employ both analysis and deliberation and argued that it is appropriate for them to do so. Among the reasons identified for using deliberation are that the most useful type of analysis often is not self-evident and is best determined through dialogue involving both the potential producers and the users of the analysis, and that judgment is inevitably involved in finding the meaning of analytic findings and uncertainties for specific decisions, particularly when the decisions must be made against multiple objectives. The report defined the challenge for public policy as one of finding procedures (called *analytic-deliberative* processes in the report) that appropriately integrate the two methods. In an effective analytic-deliberative decision process, those involved in making a decision determine the kinds of analysis they need, see that the analysis is conducted as needed, and deliberate on the choices they face, informed by the analysis and discussion of its strengths and limitations.

A central point of *Understanding Risk* was that even in such enterprises as environmental risk assessment, which are commonly seen as relying almost completely on analysis, the need for deliberation is critical. Expenditures on analysis can have little practical value if the analysis is not directed to the most important questions for decision makers. Deliberation is needed to ensure that government procures the right science for the pur-

pose. Deliberation is also critical because a single set of scientific findings can have various implications for policy, depending on judgments about matters that analysis alone cannot resolve, such as how much weight to give to the different outcomes of a policy choice that has multiple consequences and how to act in the face of gaps and uncertainties in available knowledge. Deliberation is needed to give due consideration to the possible meanings of what is and is not known. So even in very analysis-heavy areas of policy, the value of analysis ultimately depends on the quality of the deliberation that shapes and interprets the analysis.

Research policy presents a different situation from environmental and health policy with respect to the roles of analysis and deliberation. The value of deliberation is well established for making decisions about scientific research portfolios, as reflected in the careful efforts that research agencies such as the National Science Foundation (NSF) and the National Institutes of Health (NIH) make to devise and reevaluate their peer review and advisory processes. However, the value of analysis, especially that grounded in the use of quantitative measures, remains in dispute. Debate also continues about whether use of analytical methods has contributed to improved science policy decision making or has been dysfunctional (Perrin, 1998; Radin, 2000; Feller, 2002; Weingart, 2005). As already noted, the prevailing view in the scientific community emphasizes expert peer review as the most effective available method.

Reframing the debate along the lines suggested by *Understanding Risk*, that is, in terms of the appropriate roles of analysis and deliberation, may help to find optimal ways to use both sources of information. We begin by noting that all policy decisions in a democracy are ultimately deliberative. The issue is not whether to replace deliberation with analysis in making decisions, because decisions will continue to be deliberative. The issues are whether there are useful roles for analysis in a deliberative decision process and, if so, how the use and interpretation of analysis should be organized (and by whom) in research policy making. Thus, it is useful to focus attention on a set of empirical questions such as these:

- Can deliberations about the past progress of scientific fields and the best way to shape research portfolios be better informed by the use of appropriate analytic methods?
 - If so, which analytical tools hold promise for better informing judgments about behavioral and social research on aging?
 - What institutional structures and procedures are effective for selecting, shaping, and interpreting analysis to inform research policy choices?
 - How do different structures and procedures for analytic deliberation affect the distribution of decision-making influence and authority among researchers, research managers, and representatives of society?

In keeping with the tenor of mainstream conclusions of the academic research community that is the primary performer of BSR-funded research, we find it convenient to start with the judgment of the Committee on Science, Engineering, and Public Policy (1999a) report on evaluating research programs, that expert judgment is the best means of evaluating research. We further accept the widespread assessment found in the bibliometric literature that there is no single approach that will always work best and therefore that it makes sense to develop a toolbox of methods, both analytic and deliberative, for informing judgment (e.g., Grupp and Mogege, 2004). Different analytic tools might have value for different assessment purposes. They might be useful for measuring research results, organizing information brought to bear by applying other analytic tools (e.g., to arrive at numerical weights for different kinds of information), or helping to structure deliberative processes. Thus, it is appropriate to ask both about the validity of particular measures or indicators for particular purposes and about how such measures might add value to a deliberative, judgment-based process.

For convenience, we divide the following discussion into methods that are primarily analytical, those that are primarily deliberative, and those that combine both strategies. On the basis of an initial review of a larger set of decision-making techniques, we have selected three analytical approaches as most applicable to the needs for prospective and retrospective assessment as defined by BSR: Bibliometric analysis of the results of research and the connections among research efforts, reputational studies (such as can be obtained by surveying the members of research communities), and decision analysis.¹ We also discuss peer evaluation procedures, usually a purely deliberative method. Finally, we turn to analytic-deliberative approaches. A familiar one in the context of the NIH is the Consensus Development Conference, which combines analysis and deliberation but has not been adapted for making research policy decisions. We also discuss one ongoing effort in the NRC to employ an analytic-deliberative approach to a problem of comparing research in different fields, in this case, energy research.

ANALYTICAL METHODS

As noted in Chapter 4, comparative analysis of scientific progress across fields presents major challenges. The uneven pace and seemingly unpredictable paths of scientific progress and of its application to practical problems make it hard to get unambiguous meaning from even the most systematic analysis of past events in a field. Comparisons across fields are even more difficult because the paths toward progress and the barriers to it may vary systematically from one field to another. These are among the reasons that scientists and science managers have at times resisted the use of analytical techniques, especially quantitative ones, for assessing science. In addition,

there is the possibility that quantitative methods may be applied in automatic ways that exclude the judgment of the people who know the science best. In the discussion that follows, we presume that the value of analyses is not to replace judgment, but to inform it. We consider the potential roles of analytic techniques in that light.

Bibliometric Analysis

The term *scientometrics* broadly relates to the generation, analysis, and interpretation of quantitative measures of science and technology. As described by van Raan (1988b:1), the field is based on the use of “mathematical, statistical, and data-analytical methods and techniques for gathering, handling, interpreting, and predicting a variety of features of the science and technology enterprise, such as performance, development, dynamics.” Bibliometrics, the quantitative study of patterns of published scientific output and their use (e.g., citations), is the subset of scientometrics that is our primary focus of attention.²

Bibliometric and other scientometric methods were developed originally for exploring the workings of the scientific enterprise, that is, as descriptive and analytical tools, not as evaluative or predictive ones (Thackray, 1978; Godin, 2002). Their descriptive accuracy was originally validated against expert opinion. Scientometric researchers believe that a better quantitative understanding of scientific processes is needed in order to build and validate theories in the sociology of knowledge (e.g., van Raan, 2004). The distinction between the descriptive uses of bibliometrics to understand the working of science and the evaluative uses to assess performance (van Leeuwen, 2004) is important because the strengths and weaknesses of any quantitative approach, and its value to its users, depend on the questions being posed and the use to which the technique is put.

Measurement of publications and citations can be used to describe the activities of a nation, an institution, a research group, or an individual; the dynamics of fields of science that can be specified in bibliometric terms (e.g., by their leading journals or by keywords that can be found in the titles or abstracts of publications); and the relationships between and among specified fields. It can be used to build and test theories of the content and structure of science (e.g., Price, 1963), to demonstrate the contribution of publicly funded science to technological innovation (e.g., Narin et al., 1997), to highlight “hot” areas of science or hot researchers, or to track the import and export of ideas among fields.³ When bibliometric measures are treated as outputs, they can be combined with input measures, such as expenditures or personnel complements, to compare the past performance of research institutes, departments, and the like, or of fields, subfields, and

disciplines. In this use, they have the advantage of making different things comparable on the same scale.

One potentially valuable contribution of bibliometrics to the assessment of scientific fields is that it makes possible the assessment of the import and export of ideas between fields by following cross-citation patterns (van Leeuwen and Tijssen, 2000). By identifying the authors of articles published in or cited by a diverse set of journals, it is possible in principle to identify patterns of scientific collaboration across fields. It also is possible, by examining the scholarly profiles of the collaborators or their institutions or both, to assess whether particular established or newly emerging research fields are attracting the best and brightest of a nation's current and future scientists (Glanzel and Schubert, 2004; Morillo et al., 2001). Bibliometric data might also be useful for discerning and offsetting observed tendencies of proposal review panels to discriminate against "crossdisciplinary proposals that lack an established peer group" (Porter and Rossini, 1985:38).⁴

However, bibliometric measures have shortcomings as a guide to evaluative use and research decision making by mission agencies. Bibliometrics emphasizes publications in peer-reviewed journals. It does not account for practical applications that may be of value to research sponsors, research performers, and society. It provides no place for nonacademics to apply their values in gauging the societal importance of research findings. As usually implemented, it advantages journal authors over book authors or others whose works are not in major databases (Lamont and Mallard, 2005), and it favors quantitative work (which is more likely to appear in journals than books) and authors who speak to narrower and academic audiences (Clements et al., 1995, looking at sociology, in Lamont and Mallard, 2005). It favors types of research that suit high-impact journals over other types of research, such as clinical and application-based research (Kaiser, 2006). And it may overvalue scientific outputs that are frequently cited because they are controversial or wrong. Many of these shortcomings can be alleviated to a degree by careful research design, but however well this is done, the evaluative meaning of bibliometric comparisons requires interpretation, as we discuss below.

To move beyond a general review of the strengths and weaknesses of bibliometric techniques as a means of setting research priorities, we were briefed by Anthony F.J. van Raan, a leading developer and analyst of bibliometric techniques, on what one might learn from those techniques; we then commissioned a pilot study designed to determine whether it was possible to map the direction of behavioral and social science research in aging using bibliometric indicators. Committee members, whose expertise extends across (and beyond) the behavioral and social science domain of BSR's program, specified keywords intended to define certain areas of research on aging of programmatic concern to BSR. For each area, committee members

supplied an initial list of core journals in which research containing these words was likely to be published. Ed Noyons, a distinguished bibliometrician and specialist in bibliometric mapping from the University of Leiden, Netherlands, was commissioned to conduct a fuller search of bibliometric citations based on these keywords and to develop bibliometric maps of relationships between and among research clusters, journals, and authors.

The pilot study quickly revealed that the basic outputs of the exercise, such as the size of the corpus of work in a field and the boundaries of the field (e.g., which key articles are and are not included) were quite sensitive to the choice of keywords. The pilot study strongly suggested that if bibliometric indicators are to be used for research assessment, considerable reliance must be placed on the subject-matter experts to guide and review the work of the specialists in bibliometrics who will perform the actual studies. Several iterations of generation and analysis of data will probably be needed before the assigned experts are satisfied with the output. The reliability of this method, that is, the extent to which different experts' lists of keywords would yield similar results, is unknown. Thus, the meaning of analyses that are sensitive to expert judgment on the input end is likely to be open to different interpretations by experts who have different views of the research area in question. These concerns are likely to be most serious when bibliometric analysis is used to assess the dynamics of emerging research fields that lack established publication outlets or generally shared terminology.

Reputational Studies

Surveys and interviews have often been used to solicit the views of representative samples of scientific communities about issues on which judgments are to be made. An example is the periodic surveys the NRC has organized to assess research doctorate programs in American universities (e.g., National Research Council, 1995b, 2003). The reputational approach has the advantages that, unlike informal peer-review discussions that draw on reviewers' understandings of the reputations of researchers and research fields, it is systematic, it can be used continually over time, and its methods can be made transparent. The approach also has significant validity problems for its usual purpose, which is to compare entities that are presumed to be of the same type (such as university departments of psychology or economics). The problems include biases that may be introduced by relying on reputation (e.g., sensitivity to name recognition effects driven by the size of the research unit or the presence of a single well-known individual) and the difficulties of comparability among entities that may have the same names but are quite different in composition or objectives. In addition, the nature of the entities being compared can change over time, as, for example, when

taxonomies of fields become outmoded (see National Research Council, 2003, for further discussion).

Reputational approaches have additional limitations for the task of concern in this study, making comparisons across different scientific fields or subfields. Most fundamental is that there is no single scientific community that can be surveyed to get meaningful information. Very few individuals, if any, are equally well informed about each of the fields to be compared, so that sampling techniques create a difficult, perhaps insoluble, dilemma. It is possible to create an acceptable representative sample of researchers across the broad area in which comparisons are to be made (behavioral and social research on aging), but such a sample will include many respondents who are well informed about their own parts of this broad field but not about other parts. Alternatively, it is possible to create acceptable representative samples for each of the narrower fields to be compared, but this procedure will reproduce the problem that led to this study in the first place: the possibility that different standards of quality are being used in different subfields, making community judgments noncomparable across fields. We have been unable to identify a way out of this dilemma.

We do not see value in reputational studies for making comparisons of different fields without a prior demonstration that there is a valid method of eliciting comparable judgments. Value may be gained by systematically eliciting judgments of research progress from samples of narrow research communities, especially if there may be differences in judgments within the field (e.g., between younger scholars and the ones most likely to be placed on deliberative peer review groups). However, surveys should not replace judgment, and research managers need to judge whether the potential knowledge to be gained from adding a survey to judgment is worth the incremental cost of survey research. Our judgment is that it will be worthwhile only in special cases.

Decision Analysis

The above analytical methods all inform judgment by providing decision-relevant information that decision participants would not otherwise have. Decision analysis, by contrast, provides a set of techniques that can be used to organize and structure deliberation.

Decision-analytic techniques have not been given much attention in science policy, and, when proposed as decision aids, they have often met stiff resistance from scientists (Fischhoff, 2000; Arkes, 2003). We see these techniques as worthy of renewed attention because they have proved useful for assisting choices in other practical contexts in which (a) decisions are complex, (b) decisions have consequences for multiple important outcomes, (c) considerable uncertainty exists about how each choice will affect the

outcomes, and (d) opinions diverge about the relative value of the outcomes. For example, these techniques have been used to help design safety features in complex technologies, to assess the environmental and public health risks of chemicals, and to inform decisions about the siting of hazardous waste facilities. Decision-analytic techniques help clarify and allow for separate consideration of the key elements of a decision, particularly the relationships between actions and their various consequences, the valuation of these consequences, and the relationships among the decision elements (e.g., Edwards, 1954; Behn and Vaupel, 1982; Howard and Matheson, 1989; Pinkau and Renn, 1998; van Asselt, 2000; Jaeger et al., 2001; North and Renn, 2005).

Decision analysis offers both quantitative and qualitative methods. Quantitative techniques include benefit-cost analysis, multiattribute utility analysis (von Winterfeldt and Edwards, 1986), value-tree analysis (e.g., Keeney and Raiffa, 1976), value-of-information analysis (Raiffa, 1968), quantitative characterization of uncertainties (Morgan and Henrion, 1990), and prediction markets (Berg and Reitz, 2003). The usefulness of these techniques depends on the availability of quantitative estimates of the effects of policy choices or new scientific information on highly valued outcomes that are reasonably accurate or have estimable uncertainties. It also depends on developing some justifiable method for aggregating different kinds of outcomes. Because of the shortcomings of fundamental understanding of how research activities lead to scientific or technological progress (see Chapter 4), the continuing uncertainty or loose coupling of such progress when it occurs to the desired societal objectives, and the difficulties associated with aggregating different kinds of outcomes, these basic requirements are not currently met for research policy on behavioral and social science and aging. Thus, *we do not recommend the use of quantitative techniques of decision analysis to inform decisions about setting priorities for basic behavioral and social science research on aging.*

Qualitative techniques of decision analysis, by which we mean techniques for structuring or organizing decision problems without attempting to quantify the effects of decisions, are more modest in their objectives than the quantitative approaches, but they seem to have greater potential for assisting with priority setting in research policy. Decision analysis, used to structure choices, can make decision processes more transparent, thus contributing to accountability, by creating frameworks for examining issues, focusing deliberation on explicit evaluative criteria, and helping diverse groups understand the bases of their divergent judgments (North and Renn, 2005). It is likely that the best ways to employ decision science approaches for structuring research policy choices will have to be developed over time and adapted to meet particular needs (Fischhoff, 2000).

Here we note two approaches that may provide useful starting points

for such development. Both involve developing simple conceptual models of how research might contribute to a set of science policy objectives. One approach to doing this involves influence diagrams (Clemen, 1991). These are directed graphs in which each node represents a variable, arrows point from predictor variables to predicted variables, and the practical outcome variables that motivate research funding are prominently included (see Box 5-1).

Another approach specifies the objectives of the choice at hand and further specifies elements or contributing factors to each objective as a way to structure consideration of the available options. This approach was used in a recent NRC study (2005a) that recommended five priority areas for social and behavioral science research to improve environmental decision

BOX 5-1

Influence Diagrams of the Impacts of Scientific Research

Fischhoff (2000) suggests that the process of developing influence diagrams of the pathways from research to its scientific results and societal benefits can clarify the place of various research activities in the larger enterprise and promote more focused discussion of priorities, even if credible numbers cannot be calculated to estimate the strengths of the relationships that the arrows represent. Such discussion could systematically address such questions as whether anyone in the scientific community is receiving research support to understand each element in the influence diagram and “whether the research investments are commensurate with the opportunities” (Fischhoff, 2000:82).

As an example, Fischhoff presents an influence diagram in which the variable of central interest is the public health risks of *Cryptosporidium*. The diagram shows the roles of events in the biophysical environment (e.g., contamination of drinking water resulting from a flood), responses of individuals and organizations to the events, engineering practices (e.g., routine testing of the water), mass media coverage, and other factors. In such a diagram, various kinds of scientists can locate the points at which their research is relevant to reducing the risks.

This diagram emphasizes a practical, health-related outcome that research might help improve. Similar conceptual models might be developed for NIA’s practical goals for research, such as to “improve health and quality of life of older people” and to “reduce health disparities among older persons and populations” (National Institute on Aging, 2001); for considering other important NIA goals for research, such as to “understand healthy aging processes”; or for comparing research programs that contribute differentially to different research goals.

making in the public and private sectors. The study panel was given three criteria for selecting the top-priority areas: “the likelihood of achieving significant scientific advances, the potential value of the expected knowledge for improving decisions having important environmental implications, and the likelihood that the research would be used to improve those decisions” (National Research Council, 2005a:12). The panel decided that it could reduce problems of differing interpretations of these three broad criteria by specifying each criterion in more detail. Thus, it began by identifying factors that are likely to act as means to the ends implied by each criterion. For example, the panel members agreed to rate potential science priorities highly on the criterion of likelihood of achieving significant scientific advances if the following factors were judged to apply (p. 15):

- The research community is ready and able to conduct the research (e.g., concepts, methods, and data are available but not yet adequately applied in this area).
- Successful research would provide new frameworks for thinking or sources of understanding (e.g., data, methods) that could lead to advances in environmental decision making over time.
- Successful research would overcome or reduce gaps in knowledge or skill that now inhibit opportunities for improved environmental decisions in a given context.

Each panel member agreed to consider how each of the contributing factors applied to each of the potential science priorities. The panel then engaged in a discussion of each of the suggested priorities in light of the criteria and the contributing factors to each, aimed at reaching consensus on a short list of recommended science priorities.

BSR might develop a similar list of dimensions or types of scientific progress that might be made in the research fields the office supports to use in priority-setting discussions of its advisory board or other appropriate deliberative bodies. For example, research fields might be judged to be making progress along such dimensions as (Lamont, 2004):

- generativity or intellectual productivity (leading to new discoveries and theories);
- growth (e.g., attracting students and researchers, creating journals and societies, gaining funding);
- range (investigating an increasing scope of issues);
- theory development (linking a widening scope of issues within a shared conceptual framework; developing and testing hypotheses about phenomena within a common framework);

- interdisciplinarity (engaging questions raised by or of interest to other fields; framing issues that integrate previously separate fields);
- attraction (gaining the attention of researchers in other fields);
- intellectual diffusion (developing ideas or methods that are used in other fields); and
- diffusion to practice (dissemination of scientific information to potential users in fields of policy, business, law, medical practice, etc.).

BSR might sponsor a small series of organized discussions in which researchers working in different parts of the program's research portfolio would propose working lists of important scientific outcomes of BSR-sponsored research, such as those above. Such discussions would be used to generate a working list of scientific objectives for BSR-sponsored research.

On the basis of these discussions, BSR might hold further exploratory exercises to identify possible contributing factors to the key dimensions of scientific progress they identify. Such exercises might make possible more nuanced, focused, and transparent discussions about how different elements of the BSR research portfolio contribute to the institute's scientific objectives. We do not recommend that BSR attempt at this time to develop quantitative measures of these dimensions that can be used to summarize the progress of different scientific fields. Any such measures will need considerable development and validation if they are to become useful, and, regardless of how much validation work is done, we emphasize that measures of the dimensions of scientific progress should be used as inputs to deliberative processes, not as replacements for them. This is because priority setting inevitably requires a weighing of the various dimensions of progress and the various program goals.

To help structure consideration of how BSR-supported research activities may contribute to the practical goals outlined in the NIA Strategic Plan, the BSR Program might convene diverse groups of scientists and potential users and beneficiaries of the research in exercises to create influence diagrams or other simple models of the ways in which BSR research might contribute to these practical goals. The models could be used to focus deliberations about how different research activities fit into the BSR Program's objectives and where shifts in research emphasis might be justified in terms of these objectives. Again, we see these simple models as potentially useful to focus and inform deliberations about the program's research portfolio, not as a step toward developing quantitative algorithms that would take the place of deliberation.

Subsequent to developing such exercises to elaborate the practical and scientific objectives of BSR-sponsored research, the program should consider experimenting with exercises in which groups representing the producers

and various important potential users of its research deliberate together about how the BSR research portfolio can better advance the office's scientific and practical objectives. Such exercises might adapt the procedures used in the NIH Consensus Development Conference approach (see below). Such deliberations should proceed from the explicit recognition that research in different fields may be justified appropriately on different grounds. Because the objectives of BSR are widely shared across NIA, the decision-analytic exercises suggested here may be useful across the institute and deserve NIA-wide support.

As suggested by outcomes generated by various Foresight undertakings (described below), such structured deliberations may have different outcomes depending on how various constituencies are represented in the processes and what roles NIH staff and other public officials have in the process. Therefore, the processes by which they are organized should also be studied. We recommend that BSR support a series of structured deliberations involving groups that are diverse in terms of constituency to identify ways of constituting and instructing these groups that arrive at a consistent consensus that is defensible to both working scientists and agency officials. We emphasize that the value of all these decision analytic approaches is as inputs to decision making, to make consideration more systematic and transparent, not as substitutes for careful deliberation.

DELIBERATIVE METHODS

The best known deliberative approach for assessing research is peer review. Peer review panels, study sections, advisory committees, and visiting committees all engage groups of experts, usually researchers, in deliberations about science policy choices. They typically meet in person and, through discussion, arrive at collective judgments about the quality of research proposals or programs that are used as advisory input by research managers. Peer review panels typically do not rely in any explicit way on scientometric or other analytical methods (see Bornmann and Daniel, 2005). Still, there are methods to peer review. These typically involve procedures to ensure that review groups represent the full range of relevant expertise, are balanced with respect to viewpoints on matters to be deliberated, are independent of undue pressures from outside influences, and do not embody conflicts of interest, as well as procedures for review and oversight of the composition of review groups and sometimes also of their reports. Such methods are often recorded in the procedural guidelines of federal agencies, the NRC, and other organizations that routinely organize peer review groups. Decision makers in agencies may be given varying levels of discretion with regard to deviating from the collective judgment of peer review panels.

Most of the research on peer review processes concerns their use to

compare research proposals prospectively in a single field (e.g., Chubin and Hackett, 1990). Depending on the review process, the fields may vary from narrow in content to quite diverse and interdisciplinary. Although the copious literature on peer review contains numerous examples and personal observations on how the system operates in specific cases, we have found no systematic research on peer review for making comparisons among scientific fields, a type of assessment that raises issues somewhat different from those central for assessing individual research proposals. Thus, research on peer review processes must be interpreted carefully to draw conclusions about how such processes might best be used for comparing fields.

Early studies of peer review in the natural sciences show that success in obtaining funding was associated with bibliometric indicators of the quality of the investigators' past work, such as the numbers of past publications and of citations to those publications, but not to other characteristics (see Lamont and Mallard, 2005, for a review; also Campanario, 1998a, 1998b; Blank, 1991; Wessely, 1996). Other studies point to consensus among natural sciences as to the concept of quality (Dirk, 1999). These findings were widely interpreted as supporting the validity of peer review. In the social sciences, however, consensus among reviewers about quality has been seen as the exception rather than the rule (Cole, 1983; Hargens, 1987), perhaps reflecting the existence of competing standards of quality (Mallard et al., 2005). For example, a study of 12 review panels sponsored by 5 funding organizations found that reviewers in the social sciences and humanities operate with a variety of concepts of the originality of research and that some of these pertain to nonintellectual characteristics of the investigator (e.g., risk-taking, integrity) (Guetzkow et al., 2004). Reviewers also differed in the importance they placed on the potential social impact of the research vis-à-vis the intellectual quality of the scholarship and in the rationales they favored in arguing for or against supporting proposals. These differences were related to the experts' fields (e.g., social science versus humanities) and to the priorities evident in the reviewers' own research. They were also strongly influenced by the instructions that funding agencies gave to their reviewers (Mallard et al., 2005). These findings suggest that a funding organization that clearly defines its own criteria for evaluating research can convene review panels that will apply those criteria, but that without such definition, social science review panels may use inconsistent evaluative criteria.

There is some evidence to suggest that peer review disadvantages interdisciplinary research. Specifically, there is evidence that reviewers tend to favor research that belongs to their own field or school of thought (Porter and Rossini, 1985; Travis and Collins, 1991) and that follows established paths and is therefore low risk (Langfeldt, 2001). Laudel (2006) suggests, however, that such biases can be overcome by creating review panels like

those that review the German Sonderforschungsbereiche (SFB). The SFBs are interdisciplinary consortia of research groups that come up for funding renewal every 3-4 years. Their review groups are interdisciplinary and stable, and they interact in a deliberative manner over time with the SFBs, thus increasing understanding within the review panel and between the panel and the research group (see also Lamont et al., 2006).

As noted, although peer review is frequently used for assessing the relative progress of research fields and for setting priorities among them, little research exists on these efforts. For example, the NRC frequently convenes groups of experts to recommend research priorities in scientific disciplines or in interdisciplinary areas. However, these panels rarely report in detail on how they selected or applied evaluative criteria (an exception is discussed in the next section). Some commentators on peer review for comparative assessment are quite cynical about the strategic use of appointment to a peer review panel to promote reviewers' own organizational interests (e.g., Stigler, 1993; Rhoades, 2002). We have found no comparative studies of peer review processes for priority setting that examine how the structure or process of their deliberations affects the results. Similarly, we have found no studies that show how peer reviewers or peer review groups deal with the need to compare research activities that have different objectives or with the existence of diverse perspectives within review groups on the relative importance of these objectives.

Existing research on peer review thus raises but does not resolve several additional issues that we think are important for judging the progress and potential of research fields. One is whether expert review processes tend to exclude breakthrough innovations. Limited evidence can be found on both sides of this debate (Langfeldt, 2001; Rinia et al., 2001). Another is how peer review groups can deal with the different meanings of creativity in different fields.

Another important issue concerns who should be involved in the deliberations that assess scientific progress and set research directions. Scientific peer review processes by definition assume that only experts in the relevant scientific fields (i.e., the peers of the researchers) are competent to participate in deliberative review processes. This assumption has been called into question when the science is interdisciplinary (a situation that can greatly diminish the availability of true peers) and when the research is being funded for its potential practical value as well as for its potential contribution to knowledge for its own sake.

When research is being supported in part because of its potential practical value, it is often argued that the research agenda should be influenced by broadly based deliberations of groups that include both producers and potential users of the research (e.g., Committee on Science, Engineering,

and Public Policy, 1999a; National Research Council, 1996, 2005a; Renn et al., 1996). The argument is based not only on reference to principles of democracy, but also on the claim that more competent and decision-relevant choices are made when decision-making bodies have this kind of mixed representation. The approach has been employed fairly extensively in environmental and energy policy arenas, and sometimes in making decisions about basic research. It has sometimes been used in panels with a narrow purview, such as for reviewing research proposals, for which benefits have also been claimed from the inclusion of user representatives. Thus, deliberative processes that include both producers and the various kinds of users or beneficiaries of projected research deserve serious consideration by BSR in setting research directions for areas that are suspected of needing improvement in terms of their production of useful knowledge.

ANALYTIC-DELIBERATIVE METHODS

Analytic-deliberative methods are those in which judgment is based in part on information from scientific theory and data or other systematic sources of knowledge. The idea of expert judgment informed by quantitative data has several relevant exemplars, in which experts interpret quantitative measures and evaluate their import for a decision at hand. Medical diagnosis and treatment provide good examples. Physicians, acting alone or in multispecialty groups, consider available test results, the patient's reported symptoms and observable condition, and other quantitative and qualitative information before making their judgments about the correct diagnosis and the appropriate course of treatment. They monitor these same sources of information to evaluate the success of the treatment and to consider changes in it. They can make better judgments with the right quantitative test results than without them, but they use judgment in interpreting the data. Trial juries also deliberate on information that includes the results of analyses of evidence and sometimes the conflicting interpretations of the evidence by experts. As the example of trial juries suggests, it is possible to conduct useful analytic-deliberative processes in which some of the participants are not experts in the scientific issues being considered. In fact, many of the participatory processes noted above, such as those employed in environmental policy, include important roles for nonexperts in guiding and interpreting scientific analyses.

Analytic-deliberative methods in public policy are normally used by groups of people, sometimes consisting only of experts in a field, and sometimes also involving science managers or representatives of groups that might be affected by the decisions being considered. Below, we consider three examples that may be relevant to the needs of BSR.

NRC Comparative Assessment of Fields of Energy Research

The NRC's Board on Energy and Environmental Systems has produced a series of studies that compare disparate areas of energy research supported by the U.S. Department of Energy (DOE) to judge the benefits that have been gained from past research and that might flow from future research in order to assess the past performance of research areas and inform judgments about the relative priority that should be given to future research in these areas. Although energy research is in many respects quite different from behavioral and social research on aging, it is similar in certain respects that may make the energy case instructive: the research comes from disparate fields and draws on different expertise, its potential benefits are both scientific and practical, and the practical benefits are both economic and noneconomic in nature (e.g., national security, in the case of energy policy). Thus, in neither field is it easy to find expert reviewers who are competent across the range of substantive areas to be reviewed, and in neither field is there a satisfactory common metric for comparing research progress.

The committees working under the board on this effort have developed an analytical matrix designed to meet three criteria: simplicity but flexibility, transparency to decision makers, and consistency (in the sense of allowing analysis of different fields of research within a common category system) (National Research Council, 2001a, 2005c; Fri, 2004). The retrospective assessment (National Research Council, 2001a) of the benefits and costs of research used a two-dimensional matrix. The rows of the matrix distinguish three kinds of benefits and costs defined by the objectives of the DOE research effort: economic, environmental, and security. The columns reflect the certainty of the benefits. They distinguish "realized benefits," for which the technology is developed and economic and policy conditions are favorable for commercialization, "options benefits," which refer to technologies that are developed but for which economic or policy conditions are not now favorable, and "knowledge benefits," defined as "economic, environmental, or security net benefits that flow from technology for which R&D has not been completed or that will not be commercialized."

The 2001 report provides considerable detail on the kinds of benefits and costs that belong in each category and on how those benefits were to be estimated (National Research Council, 2001a:Appendix D). It emphasizes methodological considerations, such as the need to assess net benefit and the need to rely on data from sources independent of the research sponsor. The committee emphasized the need to consider all types of benefits, not only the economic ones, which are the most easily quantified. It made an explicit decision not to try to reduce each type of benefits to a dollar metric for comparisons. It collected information on the costs and benefits of the selected research programs from program managers and comments from

industry and public interest groups. Thus, the committee collected the available analytic information and organized it around the cells of the matrix, but it ultimately relied on deliberative processes to reach its conclusions (National Research Council, 2001a).

The Committee on Prospective Benefits of DOE's Energy Efficiency and Fossil Energy R&D Programs prepared a prospective assessment (National Research Council, 2005c) using a modified analytic matrix that retained the distinction among the three objectives of the R&D programs. It considered the probability of the program achieving its goals of producing new technologies and the conditional probability of market acceptance of those technologies, evaluating these outcomes in relation to three scenarios of possible energy futures.

It would be possible to develop an analogous approach for assessing behavioral and social research on aging. BSR could develop a simple but flexible evaluation methodology that is transparent and that could be applied consistently across fields. NIA strategic planning documents could specify the key objectives of BSR research. Retrospective analyses could consider a matrix of results that assessed realized benefits (e.g., to health and well-being), options benefits (e.g., development of techniques and procedures in health care), and knowledge benefits from each field in relation to the NIA research objectives. Knowledge benefits include not only knowledge that is applicable to technology or health care, but also improved basic understanding of processes of aging even if that knowledge has no foreseeable application. Prospective analyses would involve judgments of the likelihood that research investments would yield knowledge, options, and realized benefits of the types desired by BSR and NIA.

Foresight Techniques

Foresight as a technique for aiding science policy decisions has been defined as “the process involved in systematically attempting to look into the longer-term future of science, technology, the economy, environment and society with the aim of identifying the areas of strategic research and the emerging generic technologies likely to yield the greatest economic and social benefits” (Martin, 1996, p. 158). The approach is predicated on the beliefs that there are many possible futures, and that “the choices made today can shape or even create the future” (p. 159). Foresight approaches emphasize consultative processes among relevant stakeholders, with extensive provision for feedback among participants. A variety of techniques are employed to elicit projections of future trends and opportunities. These include creation of scenarios, trend analysis, Delphi techniques, technology roadmapping, among others. Foresight differs from the use made of advisory groups by federal agencies in the United States to project or recommend

future trends and opportunities in science in at least the following ways: it systematically engages a more diverse set of stakeholders in a single exercise; it employs specific techniques to structure future possibilities; and it incorporates iterative processes in which participants may modify their projections in light of information garnered about projections of other participants.

Foresight is a well-established approach for assessing prospective developments in science in several European countries, Canada, Australia, and Japan (Martin, 1996). For example, a recent review of the United Kingdom's Foresight Programme, launched in 2002, concludes that "the Programme has achieved its objectives of identifying ways in which future science and technology could address future challenges for society and identifying potential opportunities. It has succeeded in being regarded as a neutral interdisciplinary space in which forward thinking on science-based issues can take place" (Policy Research in Engineering, Science and Technology, 2006:3).

Selected consideration and use of the technique is evident among U.S. science agencies (National Academy of Public Administration, 1999). However, it has been used less frequently than standing or specially constituted advisory panels that do not employ structured Foresight techniques. Selected advisory committees and external study commissions across agencies may have considered or used variations of Foresight. Several reasons may be adduced for the comparatively limited formal use of Foresight techniques in assessing and projecting future scientific trends in the United States. One is its association with the political imbroglios that led to the demise of the Office of Technology Assessment. Although we have not attempted to evaluate past experiences or current usage of Foresight methods, we note their relevance for possible adaptation to the needs of BSR for improved methods for informing science policy decisions.

NIH Consensus Development Conference Model

The Consensus Development Conference is a familiar analytic-deliberative process in NIH. This model, which has been used more than 120 times since 1977, follows a carefully thought out rationale and set of procedures (for a detailed description, see <http://consensus.nih.gov/ABOUTCDP.htm>). It is used to produce State of the Science Statements, which summarize available knowledge on controversial issues in medicine of importance to health care providers, patients, and the public. It is also used to produce Consensus Statements, which address issues of medical safety and efficacy, may go into economic, social, legal, and ethical issues, and may include recommendations.

Consensus development conferences are deliberative in that the appointed panels discuss the implications of available scientific information

for medical practice and related issues and seek a consensus that reflects a collective judgment. They are different from the usual scientific peer review panels in that the membership is not restricted to scientists. They are analytic-deliberative because they rely not solely on judgment, but on systematic efforts to review the scientific literature and gather information from experts on the medical technology or treatment in question (analyses), and because the experts respond to inquiries from the panel and engage in discussion with it, thus closing the circle between analysis and deliberation in ways that can potentially change both processes. The consensus development process includes various safeguards of the independence and credibility of the panels, whose members are screened for bias and conflict of interest and deliberate in executive session to protect their independence from outside influence. The consensus statements are widely disseminated by NIH, but they are not government documents. They are statements by the panel, and their credibility flows from the reputations of the panel members and the procedures for ensuring that the panel is balanced, well informed, and independent.

Consensus panels are notable for their breadth of participation. They are chaired by “a knowledgeable and prestigious person in the field of medical science under consideration” who is not “identified with an advocacy position on the conference topic.” They include research investigators in the field, health professionals who use the technology in question, methodologists, and “public representatives, such as ethicists, lawyers, theologians, economists, public interest group or voluntary health association representatives, consumers, and patients.” Members are selected for their ability to weigh evidence and to do collaborative work, as well as for their absence of identification with advocacy positions or financial interest related to the conference topic.

The consensus development model has been used in NIH for providing advice on a variety of policy-related topics, but not in the area of research policy. In principle, though, elements of this model could be included in an analytic-deliberative process for advising on research policy in BSR or more broadly in NIA. To do this, several issues would need to be confronted:

- Who would be represented? For example, how broad should the participation be beyond the research community? In particular, what roles should various beneficiaries of research, from health care professionals to patients, have in advising on NIA research priorities?
- How would analysis be organized to support deliberation? Given the limitations of all the analytical approaches available in research policy, attention would have to be given to ensuring that the results of bibliometric or other methods of analysis are presented as data to be interpreted judiciously,

much as data from medical research are. The process would have to take into account the fact that the evidence on science policy choices is usually of lower quality than the evidence on medical treatments.

- How would results from a research policy consensus conference feed into institute decisions? This raises the same issues of research managers' levels of discretion and of the balance of influence and power between research managers and others that arises with ordinary scientific peer review. With nonresearchers at the table in a consensus conference setting, these issues take on a different tone.
- What are advantages and disadvantages of this model compared with current, more purely deliberative, peer review and advisory processes?

As these examples and experience in other areas of public policy decision making suggest, processes that incorporate relevant analytic techniques and information into deliberations in groups that represent the range of scientific knowledge and policy perspectives needed for wise decisions can result in recommendations and decisions with several desirable properties. The recommendations and decisions can be well informed about the available evidence, systematic in consideration of the evidence from all relevant policy perspectives, accountable, and even consensual among groups representing diverse perspectives. Because well-organized analytic-deliberative processes can entrain the full range of knowledge sources and perspectives on its interpretation, they are well suited to producing these desirable results—but they do not always produce them. Although research in some fields of public policy is beginning to identify the conditions and practices that are conducive to achieving these results (e.g., National Research Council, 1996, 1999; Renn et al., 1996), similar bodies of research have not yet been developed for the use of analytic-deliberative processes in science assessment. At present, it is worthwhile to seek to adapt practices from other fields, such as those described above, while also working to improve systematic knowledge about which processes of science assessment best meet the needs of organizations like BSR.

CONCLUSIONS AND RESEARCH NEEDS

Conclusions

1. *Assessing the progress and potential of scientific fields is a complex problem of multiattribute decision making under uncertainty.* Scientific research activities have multiple objectives, including those of advancing pure science, building scientific capacity, and providing various kinds of societal benefits. Every research policy choice and every research activity will have its own profile with regard to effects on different objectives, and there is no

agreed weighting among the objectives. Consequently, judgment is required to assess the evidence regarding how science is progressing toward each objective, as well as to consider the weight to be given to progress toward each objective.

2. *None of the available analytical methods of science assessment is sufficiently valid to justify its use for assessing scientific fields or setting priorities among them. Judgment must be applied to interpret the results from these methods and discern their implications for policy choices. This situation seems unlikely to change any time soon.* Therefore, the most appropriate use of quantitative methods is as inputs to analytic-deliberative processes of decision making. Analytic methods have the advantage in principle of making it possible to account for the progress of different fields in the same units, thus supporting priority-setting decisions in an accountable way. Each of them, however, has significant practical limitations. For example, bibliometric studies provide measures of scientific activity and of the extent to which disciplines and fields influence one another. They also have well-known limitations: they emphasize publications in the periodical literature over other scientific activities, and information about publications and citations must be interpreted in terms of the importance, correctness, and mission relevance of those activities. In addition, citation measures have been criticized as being susceptible to gaming, and reputational studies have the same limitation. Surveys of scientists to elicit their judgments are unlikely to be useful for comparing different research fields because few scientists are knowledgeable across fields, and no method is available for ensuring the comparability of judgments across the potential respondents. Quantitative methods from decision analysis are not suitable for informing science policy decisions by BSR because there is insufficient basic understanding of the paths from research activities to scientific or technological progress.

3. *Choices within NIA that involve comparisons among fields of behavioral and social science research can be better informed, more systematic, more accountable, and more strongly defensible if they are informed by appropriate systematic analyses of what these fields have produced and are likely to produce.* We consider it possible to constitute expert review panels that draw on their own experiences and insights, augmented by quantitative data on the outputs, outcomes, impacts, productivity, or quality of research, to arrive at better informed and more systematically considered expert judgments about the progress and prospects of scientific fields than they could reach without quantitative data. We think that processes that organize ongoing exchanges of judgments between bodies of scientists and science managers can produce wiser decisions than processes based on either-or thinking. Although analytic techniques should not substitute for careful deliberation, deliberation informed by analysis can produce better results

than deliberation not so informed. In Chapter 6, we offer recommendations for structuring decision processes toward this end.

4. *The scientific base for conducting valid and accountable assessments of the progress of scientific fields and for supporting research policy decisions is seriously underdeveloped.* Despite the existence of a considerable body of historical case research, little systematic knowledge exists about the paths of the development of science, particularly behavioral and social science with applications to health and well-being; about the roles of government agency decisions in that progress; about the possibility of accurately measuring and assessing such progress; about the best ways to use analytic approaches to improve decision making; or about the best ways to structure decision making to take advantage of information from studies of science.

Research Needs

Several lines of research can contribute to the knowledge base needed for a social science of science policy (Marburger, 2005) that would improve science policy decision making. This research would aim to fill the above gaps in knowledge. The research effort should be broadly based to provide broader benefits and clearer knowledge about which aspects of scientific progress are general and which are domain- or discipline-specific. In addition, a broad effort may provide general lessons about advancing interdisciplinary and mission-relevant science that can flow from research sponsored by any of a number of agencies. Research is needed to achieve the following three objectives.

1. *Improving basic understanding of scientific progress and the roles of research funding agencies in promoting it.* Research is needed to examine the nature and paths of progress in science, including the roles of decisions by science agencies. To support BSR, research is needed on progress in fields of behavioral and social science related to aging.

Scientific progress is usefully understood in terms of a causal stream that roughly moves from (a) processes that structure research to (b) inputs to research to (c) scientific outputs to (d) scientific outcomes to (e) impacts on society, as these terms are defined in National Research Council (2005c) and elaborated in Chapter 4. Society closes the circle by providing inputs and structure for research, generating research questions, and in other ways. But for the purpose of evaluating the programs of science agencies, it is useful to focus on how variables earlier in the stream affect variables later in the stream. Thus, scientific progress is usually evaluated on its outcomes and impacts. Assessments of research programs must consider these consequences in light of the level of effort (e.g., processes and inputs) that went into trying to achieve them.

Research on the nature and paths of scientific progress can build basic understanding of conditions that facilitate and impede such progress. The research might include:

- Historical analyses of the evolution of scientific fields—their rise, continued fecundity, or decline—performed by or vetted by professional historians to ensure adherence to professional standards, especially in attributing causation. “Stories of discovery” or progress, as supported by BSR and other federal science agencies, while useful in putting a face to agency claims of contributing to scientific advance, are limited as tools of analysis. They are subject to selection bias that arises from examining only the successes from among the investments made by an agency or program. They also tend to highlight agency-specific contributions, without considering the importance of other sources of contributions to progress. What are needed are studies of fields that are generally considered to have been productive and of fields that are not so considered, conducted in a manner that meets professional historical standards (e.g., Nye, 1993; Kohler, 2002). These studies could usefully focus on how the processes that organize research programs and the inputs to those programs have affected scientific outputs, outcomes, and impacts.

- Advanced bibliometric analyses of the development of research fields, to provide a window into the development of research fields over time and the flows of influence among them. These studies should look at outputs in relation to measures of inputs and processes and also in relation to indicators of scientific outcomes and impacts. Particular emphasis should be placed on the cross-fertilization of research findings from one disciplinary domain to another and the emergence of new fields of knowledge. Some of the studies should consider the usefulness of bibliometric indicators of research outputs as measures of research progress, defined in terms of each of the sponsoring agency’s program goals. Such potential indicators will require careful methodological analysis to assess their validity and potential biases before they are ready for practical use, even as inputs to decisions.

- Studies of scientific progress using emerging databases of conference proceedings or other prepublication scientific outputs. In many fields, new research results are first presented in technical reports or at conferences. Data on such kinds of activity may provide earlier indicators of scientific progress than bibliometric measures.

- Analyses of research vitality or interest shown by active scientists in lines of research, focusing on research directions that are widely considered in hindsight to have been successful or unsuccessful in terms of yielding major scientific advances or societal impacts. The studies should examine the ways that the vitality of scientific fields may relate to subsequent scientific outcomes and impacts. For example, studies should be made of the

career paths of productive scientists (“stars”) in terms of their choice of research topics, the journals in which they publish, and the career paths of the graduate students they train. Such studies could test the hypothesis that progress in a field can be predicted from the quality of the researchers who are willing to allocate their time to a specific line of inquiry.

- Studies of the effects of the structure of research fields on their progress. These studies might compare the consequences for the development of scientific fields, particularly new fields, of research portfolios that emphasize large centers, database development efforts, or interactive workshops, with more traditional research portfolios emphasizing funding to individual investigators and small research groups.

Research on the roles of science agency decisions in scientific progress can help the offices and agencies that sponsor it to make decisions about how to select and train research managers and organize advisory groups so as to better promote program goals for advancing science. This research might include:

- Studies of the role of officials in science agencies in promoting scientific progress. Some of these studies might follow the example of past research done for U.S. foundations (e.g., Kohler, 1991; Rooks, 2006) that has investigated how program managers have acted as entrepreneurs who help build new research fields and as stewards of vital fields. The research might also include studies of the characteristics of effective research entrepreneurs and stewards and studies of the effects of science agencies’ practices of hiring, training, and evaluating program managers on their scientific entrepreneurship and stewardship.

- Studies of how expert advisory groups, including study sections and advisory councils, make decisions affecting scientific progress (e.g., comparing decision making in disciplinary versus interdisciplinary advisory groups; examining the effects of emphasizing explicit review criteria, such as innovativeness, on group decisions; examining how review groups consider multiple decision criteria; investigating hypotheses, such as that peer review groups generally select in favor of methodological rigor at the expense of innovation and that different advisory groups have distinct cultural differences that affect their ability to nurture scientific innovation).

- Studies of the effects of the organization of advisory groups on their success at promoting interdisciplinary and problem-focused scientific activity and ultimately at improving scientific outcomes and societal impacts. These studies might examine the roles of advisory group chairs in shaping group decision rules; the effects of the characteristics of group members individually and collectively; and the processes of training, mentoring, and

socializing advisory group members and of oversight of advisory group processes.

2. *Improving understanding of the uses of analytic techniques in making research policy decisions.* This research would support the development, trial use, and empirical investigation of a variety of quantitative measures and decision-analytic techniques for assessing the results of past research investments and setting research priorities. The studies would seek to validate analytical techniques and to determine their best uses, which may be different for different analytic techniques. The research might include:

- Studies comparing multiple indicators of research vitality, outputs, outcomes, or impacts of lines of research with each other and with the unaided judgment of experts in these areas to see whether it is possible to develop reliable and valid quantitative measures of scientific progress through a convergence of indicators and to determine whether any such measures might be useful as leading indicators that predict critical scientific outcomes or impacts.
- Comparative studies of fields that are widely judged to differ in rates of progress toward positive outcomes and impacts to see whether particular quantitative indicators or a convergence of indicators yield results consistent with expert judgment.
- Studies to assess the value of providing information developed through specific analytic techniques, such as bibliometric studies, for research priority setting. Studies using cross-citation patterns or analyses of academic and professional career trajectories of researchers and students can show whether such analyses add significantly to the decision-relevant knowledge of expert review groups and whether and how this information alters their recommendations.
- Studies of scientific impact using databases that cover citations in policy documents and the popular press, with the results examined from the perspectives of research scientists and policy makers.

Tests of ways to employ a convergence of information from different analytic methods to inform priority setting. This research might identify whether certain ways of combining information from multiple sources can contribute to more robust and reliable decision making than reliance on any single method.

3. *Improving the incorporation of techniques for analysis and systematic deliberation into advisory and decision-making procedures.* This research should explore and assess techniques for structured deliberation, some of

them incorporating information from potential quantitative indicators of scientific progress and potential, for retrospective assessment and priority setting. The research would be used to elaborate and refine deliberative methods for organizing peer review and expert advice. The research should include the following:

- Studies to develop techniques for structuring decision analysis for use in the research priority-setting tasks facing BSR. Some studies might develop influence diagrams modeling the relationships of scientific activities (processes, inputs, outputs) to BSR goals (especially outcomes and impacts) and explore the feasibility of using these to structure deliberation. The influence diagrams might be developed by outside researchers or BSR staff, in consultation with the program's advisory council. Some studies might explore ways to structure discussions within deliberative groups around the multiple goals in the NIA program plan or around lists of types of scientific outputs, outcomes (e.g., dimensions of scientific progress), and impacts. These studies might involve the use of simulated advisory groups.

- Trials of analytic techniques for informing and structuring decisions in the deliberations of actual review and advisory panels or shadow panels created for experimental purposes. Some studies might provide panels with the most relevant available quantitative indicators for their tasks and leave them a period of time during their deliberations to discuss the meaning of the indicators for the decision at hand. Resources permitting, parallel panels could serve as comparison groups. In some studies, panels would be asked to apply structured methods for considering quantitative and qualitative information about the activities in the fields to be compared in relation to explicit criteria, such as lists of BSR strategic goals or dimensions of scientific progress, or to use influence diagrams showing plausible paths from research to the achievement of desired program goals. The studies would examine the effect of the interventions on (a) panel members' reports of whether and how their thinking or their recommendations were affected; (b) indicators of decision quality, such as the number of relevant decision objectives and pathways from decisions to the achievement of objectives that are considered in the deliberations; and (c) the creation of a sufficiently explicit record of the rationale for the advisory panel's recommendations to improve accountability and allow for a better informed exchange of judgments between researchers and research managers.

- Studies to adapt existing analytic-deliberative assessment approaches, such as the NIH Consensus Development Conference model to the purposes of assessment of research areas and research priority setting in BSR. Some of these studies might incorporate the above techniques for informing and structuring decisions. Some of the studies might include nonscientists, selected to represent the perspectives of the potential users or beneficiaries

of the research, in the analytic-deliberative process. These studies could explore how adding these perspectives may affect the ways in which the advisory groups assess the benefits of research for basic understanding and for society.

- Comparative studies of advisory panels of different composition, particularly for recommending research priorities. For example, BSR, NIA, and the NIH Center for Scientific Review might vary the breadth of expertise of experts or the balance between senior and junior researchers. Such research would provide an empirical base for assessing the reliability of deliberative advice and the sensitivity of the advice to the intellectual backgrounds and practical orientations of panel members. Such experiments would also offer evidence to evaluate such claims as that panels of researchers are too conservative to support promising high-risk research or too uncritical in areas of expertise of only one or two panel members.

- Studies involving the instruction and training of advisory panel members to consider specific BSR and NIA objectives, including mission relevance, that go beyond generic considerations of the quality of proposed research.

NOTES

1. Treated as subsets of the broader methodologies covered in this report and thus omitted from specific discussion are various Foresight techniques (Irvine and Martin, 1984) and mechanisms for scoring R&D priorities.
2. Analysis of other prominent performance measures contained within the larger scope of scientometric inquiry, such as patent statistics and publication-patent relationships, are not relevant to much of BSR's research portfolio, which produces different kinds of impacts.
3. Debates about the relative contributions of theoretical and empirical approaches to scientific advance and about leader-follower relationships between them are staples in the history of science and entail issues that extend well beyond the scope of this report (see, e.g., Galison, 1999).
4. Bibliometric evidence on the social sciences, for example, consistently shows that sociologists and political scientists cite articles from economics journals more frequently than economists cite sociological or political science journals. These findings have been alternatively interpreted as indicating the greater generalizability and precision of economic modes of analysis, and thus its greater intellectual vitality, and as documenting the intellectually closed-loop, solipsistic nature of economic thinking (Laband and Pietter, 1994; MacRae and Feller, 1998; Reuter and Smith-Ready, 2002).

6

Conclusions and Recommendations

The Behavioral and Social Research (BSR) Program at the National Institute on Aging (NIA) asked the National Research Council to undertake this study to “explore methodologies for assessing the progress and vitality of areas of behavioral and social science research on aging, and to identify the factors that contribute to the likelihood of discoveries in areas of aging research.” The ultimate purpose was to “seek practicable approaches that can help research managers improve their judgments and research portfolios.”

These purposes are perennial in U.S. science policy. They are important not only to BSR, but also to many other science and technology research organizations across the federal government, to scientific communities, and to the science policy community. Our findings and recommendations are intended as a direct response to the questions posed by BSR, but we have considered this response in this larger national context based on the thinking that it may be useful to other federal agencies that provide support for scientific research.

Long-established procedures for determining research priorities and allocating research funds in federal science agencies are increasingly being questioned for several reasons.

- *Tighter funding:* In an environment of projected static or declining research budgets for other than national defense and homeland security, proposals to open up new areas of scientific inquiry, support currently dynamic fields, and support the increased costs of existing lines of research become

competitive; with increasing force, they imply reallocations of funds, including possible reductions in support for long-established fields.

- *Increased demands for accountability and documented performance* of all federal agencies, including research agencies, which are increasingly accompanied by calls for the use of quantitative performance measures.

- *A belief that existing peer review procedures are unduly conservative* in identifying or supporting transformative, interdisciplinary, and translational research that truly presses against the frontiers of science or integrates research findings with clinical applications.

- *Science program managers' efforts to justify their decisions*, especially when these involve launching new scientific initiatives with static budgets or adjusting program priorities in ways that may conflict with recommendations from scientific advisory groups.

- *Developments in analytical methods*, databases, and statistical and data mining techniques that promise better ways to assess the impacts of lines of research on knowledge and agency societal objectives.

These pressures on standard decision processes come from inside and outside federal science agencies. They reflect the desire of agency officials to be more proactive, entrepreneurial, and responsive both to the dynamics of scientific discovery and to external pressures on the agency. They also reflect the concerns of some sectors of the scientific community, including established researchers who are seeking to extend their work into new fields and newer researchers seeking to venture into relatively uncharted domains, about the rigidity and conservatism of established review mechanisms.

Multiple variations on existing procedures are being considered. These include changing the composition of review panels; changing the criteria, methods, and means by which review panels function; and changing the relative decision-making authority of review panels and agency officials. Pervading all considerations about changes to existing peer review procedures is the recurrent concern that any such changes should not detract from the workings of a national system for allocating federal funds for research that has historically been associated with the rise to preeminence of U.S. science and graduate education.

The merit-based, peer review procedures that have become traditional in NIA and many other science agencies reflect a political consensus about the societal utility of allowing the “republic of science” to rule within constraints defined by national priorities set through budgetary and regulatory processes. Merit-based, peer review procedures serve not just as sorting mechanisms to generate the best science; they also provide essential safeguards against the insertion of patronage or ideological factors into the selection of research proposals and research performers. Consideration of

changes to current peer review procedures, whatever may be their merit in terms of opening up selection processes to newer, fresher, more relevant theories and approaches, must not lose sight of the larger institutional and inherently political context of federally sponsored research. One of the key issues is how particular procedural changes might alter the distribution of power and influence between scientists and agency administrators and research managers.

Alternatives to peer review, although increasingly discussed, have to date been tried only by limited implementation of variations on standard practices. Thus, very little is presently known about the likely impacts of such changes on scientific performance or attainment of agency mission objectives. All proposals for change thus should be considered as hypotheses that improved outcomes will follow upon their adoption; logically and necessarily then, implementation of such changes should be accompanied by careful and systematic evaluation to determine whether or not they produce the desired results.

CONCLUSIONS

1. *The scientific base for conducting valid and accountable assessments of the progress of scientific fields and for supporting research policy decisions is seriously underdeveloped.* Despite the existence of a considerable body of historical case research, little systematic knowledge exists about the paths of the development of science, particularly behavioral and social science with applications to health and well-being; about the roles of government agency decisions in that progress; about the possibility of accurately measuring and assessing such progress; about the best ways to use analytic approaches to improve decision making; or about the best ways to structure decision making to take advantage of information from studies of science.

2. *No theory exists that can reliably predict which research activities are most likely to lead to scientific advances or to societal benefit.* It is for this reason that the case for expert judgment continues to remain persuasive. Evidence of past performance and current vitality, that is, heightened interest among scientists in a topic or line of research, are imperfect predictors of future progress. Thus, any choice to support an emerging research direction is speculative. Scientific managers can best defend such choices by developing an explicit rationale for allocating funds among established and emerging fields and for making choices, particularly involving the latter.

3. *Science produces diverse kinds of benefits by diverse mechanisms that are not well understood. Consequently, assessing the potential of scientific fields or lines of research is a complex problem of multiattribute decision making under uncertainty. Investment strategies suitable for uncertain conditions are therefore appropriate for managing the BSR portfolio.* Re-

search activities can advance science on several dimensions, which we have summarized under the broad categories of discovery, analysis, explanation, integration, and development, and they can contribute to society on several other dimensions (identifying issues, finding solutions, informing choices, and educating the society). Every research policy choice and every research activity will have its own profile with regard to how much progress it is making across these dimensions, and there is no agreed weighting of the importance of one against another. Consequently, judgment is required to assess the evidence regarding how science is progressing toward each objective, as well as to consider the weight to be given to progress toward each one. The policy implications of a finding that a line of research is or is not currently making much progress on one or more dimensions are not self-evident. Hot areas may prove in retrospect to have been fads. A field that appears unproductive may be stagnant, fallow, or pregnant. Telling which is not easy.

Given BSR's environment of complexity and uncertainty, the following investment strategies seem appropriate:

a. *Portfolio diversification strategies that involve investment in multiple fields and multiple kinds of research.* Such strategies can allow research managers to minimize the consequences of overreliance on any single presumption about what kinds of research are likely to be most productive. Diversification is also advisable in terms of the kinds of evidence relied on to make decisions about what to support. For example, when quantitative indicators and informed peer judgment suggest supporting different lines of research, it is worth considering supporting some of each.

b. *Investing where the investment is most likely to add value.* For example, although directly contributing to major discoveries remains the gold prize of federal science agencies, more indirect methods, such as supporting improvements in databases and analytic techniques, integrating knowledge across fields and levels of analysis, calling attention to underresearched questions, and facilitating the entry of new people to work on old and new research problems, can yield high scientific and societal returns. By promoting scientific analysis, integration, and development, research managers can contribute indirectly to discovery and explanation. Research managers should also consider favoring support to research organizations or in modes that have been shown to have characteristics that are likely to promote progress. And BSR managers may reasonably prefer to support research in fields that need only small investments from NIA to leverage funds from other sources or in which BSR seems the only viable sponsor for the research.

c. *Support for issue-focused interdisciplinary research.* Interdisciplinary research has significant potential to advance scientific objectives that

research management can promote, such as scientific integration and development and scientists' attention to societal objectives of science consistent with BSR's mission. Moreover, BSR has a good track record of promoting these objectives through its support of selected areas of interdisciplinary, issue-focused research.

4. *Both working extramural scientists and NIA program managers have essential perspectives to contribute to research priority setting.* For example, extramural scientists often have a keener understanding of the theoretical and methodological quality of research in their areas of expertise and of which research problems are tractable given existing data and methods. By contrast, NIA program managers may have a keener understanding of the potential for linking recent developments in disparate fields that are not yet communicating, and of the ways certain lines of research might influence policy decisions in the health sector. It follows that both groups should have roles in priority setting and that an exchange of ideas among their various perspectives can promote enlightened priority setting.

5. *None of the available analytical methods of science assessment is sufficiently valid to justify its use for assessing scientific fields or setting priorities among them. Judgment must be applied to interpret the results from these methods and discern their implications for policy choices. This situation seems unlikely to change any time soon. Although analytic techniques aimed at quantifying scientific progress can provide useful input to decision-making deliberations in BSR, they should not be used as substitutes for judgment or deliberation.*

Analytical methods for assessing scientific progress and potential are those that use "rigorous, replicable methods developed by experts" (National Research Council, 1996:20). In science priority setting, these include bibliometric analysis and decision-analytic techniques such as benefit-cost analysis. Their inadequacies as aids to decision reflect (a) uncertainties inherent in projecting the future development of scientific fields on the basis of their past performance, (b) uncertainties and unknowns concerning the relationships between measurable scientific activity and the kinds of outcomes sought by the NIA, and (c) the difficulties of comparing research activities that are likely to contribute to different Institute objectives, such as scientific understanding and societal benefit. In addition, quantitative analytical methods typically have limitations associated with data collection, reliability, validity, cost, timeliness, and acceptability, as well as the lack of knowledge about how best to combine measures of qualitatively different aspects of scientific progress.

Resistance from federal science agencies and their advisory groups to the introduction of analytic techniques into decision-making processes partly

reflects a concern that they may be misapplied or applied to the exclusion of good judgment. Indeed, the use of these methods in a reductionist, bureaucratic approach to priority setting and assessment potentially threatens the validity of research assessment and, ultimately, the vitality of the U.S. research system. Resistance to the use of quantitative methods may also reflect the possibility that the use of techniques that can be applied without relying on the judgment of scientists will alter the distribution of authority and influence among such parties as agency officials, program managers, study group panels, and individual reviewers. Acceptance or rejection of specific methods may therefore reflect matters of organizational politics as well as evaluation methodology. These are legitimate concerns.

6. *Despite the many limitations of analytic techniques for assessing science, judgments in NIA that involve comparisons among fields of behavioral and social science research can be more systematic, more accountable, and more strongly defensible if they are informed by appropriate use of systematic analytic techniques.* Although no analytic technique is sufficiently developed to replace judgment, judgment can be disciplined and enhanced by careful analysis. Analytic techniques should have two main roles: (1) to help structure the deliberations about research priorities by scientific advisory groups to BSR and by the program's decision-making bodies and (2) to help structure communication between institute officials and their scientific advisers about priority setting (e.g., by clarifying the sources of any disagreements in judgment between them).

We are saying that neither judgment nor any foreseeable analytic technique provides a gold standard for science priority setting. However, we think that wise integration of analysis and judgment may yield better results than either approach unaided by the other. We consider it possible to constitute expert review panels that draw on their own experiences and insights, augmented by quantitative data on the outputs, outcomes, impacts, productivity, or quality of research, to arrive at better informed and more systematically considered expert judgments about the progress and prospects of scientific fields than they could reach without quantitative data. A key to the effective use of analytic techniques is to embed them in deliberative processes that explicitly consider clearly specified decision objectives and that focus deliberation on the expected relationships of particular decision options to particular objectives. Structuring deliberation in this way can encourage more explicit consideration of how particular kinds of research may promote specific goals of the BSR Program and provide a way of thinking carefully about what analytic techniques do and do not offer. Structured deliberation can be usefully employed in advisory groups, in program management settings, and for structuring communication about priority setting between BSR and its advisers.

PRINCIPLES FOR PRIORITY SETTING

On the basis of our consideration of the state of knowledge about scientific assessment and priority setting and of the specific programmatic setting of BSR, we have arrived at a set of principles for priority-setting activities. These principles undergird a strategy for decision making that relies primarily on institutionalizing processes for coping with uncertain knowledge and secondarily on analytic techniques intended to reduce uncertainty and aid deliberation on matters of judgment.

The best way to make good, defensible priority-setting decisions is through processes of open, explicit dialogue that are organized to raise all the major decision-relevant issues, allow for input from all relevant perspectives, and provide for iterative discussion between researchers and science managers and for orderly reconsideration of past decisions. Such dialogue can also improve the accountability of decision making. Thus, we recommend adopting processes that can collect the best available information about the progress and prospects of the areas of science being considered; use systematic procedures to consider, interpret, and discuss the import of this information for the decisions; involve extramural scientists, agency officials, and user communities in the decision processes; encourage them to consider the information in the light of all of BSR's strategic objectives; give advisory group discussions significant weight as input to decisions within the institute; allow science managers to reallocate funds among lines of research in light of advisory group judgments and their own; and allow the institute to learn from the results of its decisions.

Three principles should guide BSR practice in setting priorities across research fields:

1. Explicitness. *Judgments about the progress and potential of scientific fields should be based on explicit consideration of them in relation to all the major scientific and societal goals of the BSR Program and all the major processes and inputs supporting progress in each field.*

BSR, like many other federal science agencies with both scientific and societal objectives, must consider disparate objectives in assessing scientific progress and setting priorities. Box 2-1 shows the most recent articulation of the research goals of NIA. Different lines of research are likely to have different profiles in terms of how much they contribute to these disparate goals. Priority setting involves consideration, either explicitly or implicitly, of how much each line of research is likely to contribute to each goal, as well as about the relative importance of the goals. Moreover, different lines of research require different inputs, and these needs must also be considered separately for each field if rational priority setting is to occur. Making such considerations more explicit and deliberating about the bases for judgments,

including disagreements about judgments, will contribute to better considered and better justified decisions and recommendations for action.

The principle of explicit consideration is based in part on research on decision making that shows the limitations of unstructured judgments about multiattribute choices and the potential benefits of explicit consideration of all the relevant objectives and the values associated with them. In addition, we think that explicit consideration will make decisions in BSR more accountable and contribute to high-quality communication between researchers and science managers. The principle should be applied to recommendations made by BSR's advisory council and other groups organized to advise on priority setting. It should also be applied to internal decision making by BSR Program managers.

2. Perspective. *Both extramural research scientists and institute program managers should be involved in assessing the progress and potential of the research fields supported by the BSR Program.*

Wise priority-setting requires both depth and breadth of view, as well as concern with both the scientific and societal goals of research. For these reasons and others, it requires input both from research scientists and program managers, as well as an appropriate and productive balance of influence and power among them. Both sets of contributors to priority-setting decisions bring valuable knowledge and insights to the process, but they have different perspectives—sometimes conflicting, but nevertheless complementary. Research scientists typically see the value of research in their own fields more easily than they see the value of research in distant fields. Compared with program managers, they typically have a deeper understanding of the science in their fields, but a narrower range of view across fields. They may also give more weight than program managers do to scientific goals relative to other agency mission goals. Program managers are frequently better able to gauge the relevance of a line of research to agency mission objectives. They are typically better positioned to observe converging or intersecting advances or trends in science that fall between or at the outer boundaries of existing disciplines, and they are manifestly better positioned to support embryonic lines of inquiry. They are well positioned to observe the extent to which fields nurture one another: to observe which mainstreams spill over their banks to contribute to the productivity of other fields and which evolve in ways that produce value only to those in the field, with limited contribution either to the broader scientific enterprise or to the BSR mission.

3. Iteration. *Priority-setting exercises should be conducted regularly, and they should include reconsideration of past decisions.*

Scientific priority-setting decisions are made under conditions of uncertainty: the larger, more discontinuous the scientific advance being sought, the higher the risk. Moreover, because of the likelihood of surprise in the history of science and of change in societal priorities, even the most judiciously considered decisions need to be revised from time to time. The process of reconsideration should include reflection on the value of the analytical information provided about scientific progress to inform those decisions and the adequacy of the methods for producing that information.

RECOMMENDATIONS

We make five recommendations for implementing the above principles to develop a stronger scientific basis for research priority setting and to strengthen the basis of priority setting in BSR both before and after research results come in. We think these recommendations can be adapted for use in other science agencies that, like BSR, have multiple objectives to advance science and benefit society.

1. The staff of the BSR Program, with the help of the program's scientific advisers, should develop an explicit list of scientific outcome and societal impact goals for the program in line with the strategic program goals of NIA. Information from the staff to advisory groups regarding the progress of program-supported research should reference these goals.

Box 2-1 lists NIA's research goals; the goals of BSR are likely to be consistent with these, but more specific and perhaps more elaborated. Chapter 5 includes a list of dimensions of scientific progress that may help BSR elaborate its list of scientific outcome goals. It also suggests procedures by which BSR might identify more specific sets of scientific objectives for the program, as well as factors believed to contribute to achieving these objectives. The BSR staff already provides useful information about research progress to the advisory council—for example, in its stories of discovery and reports on science advances. We recommend that staff frame such information at least in part in terms of explicit program goals, thus promoting a more focused discussion of the kinds of scientific progress being made in each of the lines of research BSR supports and the kinds of future advances that might be expected. Explicit consideration of program goals may lead BSR staff to develop additional information beyond what is now provided to the council.

We recognize that efforts to increase accountability by justifying program activities in terms of broader goals can easily degenerate into bureaucratic exercises aimed only at complying with administrative mandates. The intent of this recommendation is not to call for new administrative require-

ments, but rather to guide ongoing science policy deliberations toward more explicit and focused discussion of how particular activities may or may not advance program goals.

2. NIA should periodically conduct a general assessment of the BSR Program with respect to its overall adequacy for supporting the program's scientific outcome and societal impact goals.

Assessments should be conducted approximately every four years, either as part of the regular quadrennial review of the BSR Program or as a separate activity by an ad hoc advisory group organized by the NIA advisory council. General assessments should explicitly consider each program goal in relation to each aspect of the BSR Program judged to be important for achieving it (e.g., the different kinds of research activities supported and modes of support). The assessments may consider indicators of scientific output (for example, bibliometric measures of publications and citations or evidence of presentations at scientific conferences), but, as they do so, the meaning of output indicators should be discussed explicitly in light of the program's scientific outcome and societal impact goals. Outputs should not be taken as evidence in themselves of progress toward scientific or societal goals. The assessments should consider the value added by NIA investment vis-à-vis investments in the same fields by other agencies and the private sector. On the basis of these considerations, the group conducting the assessment should advise on ways to alter processes or inputs in the BSR Program if this is needed to advance the goals. Following the perspective principle, these periodic assessments should be based on input from both research scientists and program managers. Following the principle of iteration, the assessments should include reexamination of past assessments.

3. NIA should periodically conduct an area-based assessment of the BSR program that includes recommended priorities for new and continued support among the substantive areas of research included in the program. These efforts should explicitly assess and compare the past and potential contributions of research in each area receiving major BSR support with regard to each of BSR's goals for scientific outcome and societal impact and with respect to the various inputs and processes that contribute to achieving the goals.

Area-based assessments should be conducted approximately every four years, either as part of the regular NIA review of BSR or as a separate activity organized by the NIA advisory council. These assessments should consider information provided by BSR staff about progress in each area or line of research, referenced to program goals. Indicators of scientific output

in each area should be collected and reviewed. The group conducting the assessment should recommend priorities after considering the progress of each area, BSR's inputs to the area, and inputs from outside BSR that may affect the value added by BSR contributions. Following the principles of perspective and iteration, it should draw on input from both research scientists and program managers and reflect on and reexamine past assessments. It should consider and make recommendations as appropriate for each area on issues of portfolio allocation between disciplinary and interdisciplinary research; basic and applied research; high-risk and low-risk research; development of research methods, of data, and of findings; support of research centers, program projects, and individual investigators; and support of research, infrastructure, and human resources development.

In conducting area-based assessments, NIA should explicitly consider BSR's activities in each area against each of the program's scientific impact and societal outcome goals, perhaps by organizing discussions around specific, stated objectives and explicitly addressing the issue of how to combine objectives in making recommendations. We caution, however, about adopting procedures that add together individual or group judgments of how research contributes to each BSR outcome goal multiplied by a weight for each goal.

Our caution results from the judgment that the appropriate way to integrate multiple goals is not always additive. A line of research that clearly produces desired societal benefits may be highly justified even if it does little to advance basic understanding. The reverse is also true: a line of research that generates a breakthrough in basic understanding may be highly justified even if it produces no current or foreseeable practical benefits for human health and well-being. All of BSR's program goals should be considered in making area-based assessments, but the group conducting the assessment should have flexibility in how it does that. It should, however, be explicit in its assessment reports about how it addressed the multicriteria aspect of the assessment.

4. The BSR Program director should consider the area-based assessments and recommendations carefully in reallocating funds among fields. One year after completion of each area-based assessment, BSR staff should report on decisions reached and actions taken that involve priority setting among research areas and portfolio allocation within areas. The report should explicitly discuss the justification for program decisions that might seem inconsistent with the assessment's recommendations. The report should be delivered to the NIA director and the NIA advisory council.

This report should not be construed as a mechanism for tying BSR decisions tightly to the recommendations of the assessment group. There

can be good justifications for institute decisions that deviate from the recommendations of a body of scientists. The purpose of the report is to ensure that such justifications are made explicit and open to question, thus providing increased accountability in an institutional sense and promoting a continuing rational dialogue among scientists and program managers, all of whom want to make the BSR program effective and productive, focused on the program's objectives. We presume that the NIA advisory council may comment on these staff reports either informally or in subsequent formal assessments, thus maintaining the iterative process of rational deliberation that involves both science managers and extramural scientists.

5. The NIA BSR Program, together with the rest of NIA and the National Institutes of Health, as well as the National Science Foundation and other federal science agencies, should support a coordinated program of research to advance well-informed, high-quality research policy making.

This research would support the need recently articulated for a social science of science policy that would improve the knowledge base for setting priorities for scientific research (Marburger, 2005). A multiagency effort is desirable because it can yield benefits that all science agencies can share, such as general lessons about advancing interdisciplinary and mission-relevant science and improved understanding about which aspects of scientific progress are generic and which are domain- or discipline-specific. The research would aim to achieve three objectives:

a. *Improve basic understanding of scientific progress and the roles of research funding agencies in promoting it.* Research pursuing this objective would examine the nature and paths of progress in science, including the roles of decisions by science agencies. It might include historical analyses of the evolution of scientific fields that have differed in their productivity; advanced bibliometric analyses that examine scientific outputs in relation to measures of inputs and outcomes, the cross-fertilization of research findings, and the emergence of new fields of knowledge; and studies of how the structure of research fields affects their progress. Research on the roles of science agency decisions in scientific progress might include studies of the role of officials in science agencies as entrepreneurs and stewards of scientific fields; studies of how expert advisory groups, including study sections and advisory councils, make decisions affecting scientific progress; studies of the effects of the organization of advisory groups on their success at promoting interdisciplinary and problem-focused scientific activity, and ultimately at improving scientific outcomes and societal impacts. In the case of BSR, the research should focus on progress in fields of behavioral and social science related to aging.

b. *Improve understanding of the uses of quantitative decision aids in making research policy decisions.* This research should include the development, trial use, and empirical investigation of the use of quantitative measures and decision-analytic techniques as inputs to priority setting. It should not seek techniques that can supplant deliberation, because different areas of science make different kinds of progress and judgment will always be required to assess progress against multiple objectives. The research would aim to identify useful techniques and determine how to use them effectively. The research might include studies to assess the value of providing information developed through specific analytic techniques (e.g., bibliometric analysis, techniques derived from decision analysis) to the deliberations of expert review groups; studies comparing multiple indicators of research activity, outputs, outcomes, or impacts and indicators of scientific progress with each other and with unaided expert judgment to seek reliable and valid quantitative measures of scientific progress and to determine whether any of them might be useful as leading indicators; comparative quantitative studies of fields that are widely judged to differ in rates of progress to identify quantitative indicators that yield results consistent with expert judgment; tests of ways to combine information from different analytic methods; and studies of the use of techniques of decision structuring and multiattribute evaluation for guiding deliberation about research priority setting.

c. *Develop useful techniques for systematic deliberation in advisory and decision-making procedures.* Research pursuing this objective should explore and assess techniques for structured deliberation, some of them including the use of indicators of scientific progress and potential, for retrospective assessment and for priority setting. It should include studies to apply techniques for structuring deliberation to the research priority-setting tasks facing BSR, possibly using simulated advisory groups; trials of such techniques in which review and advisory panels are instructed or trained to focus their deliberations on how each research field might contribute to specified program objectives or goals, including both those related to scientific quality and to mission relevance; studies of attempts to adapt the NIH Consensus Development Conference model to research priority setting; comparative studies of advisory panels or simulated advisory panels of different composition to test hypotheses about the limitations of current review panels in recommending research priorities; and studies of the effects on decision making of instruction and training of advisory panel members to consider the full range of specific BSR and NIA objectives.

References

- Abrams, P.A.
1991 The predictive ability of peer review of grant proposals: The case of ecology and the U.S. National Science Foundation. *Social Studies of Science* 21:111-132.
- Adams, E.K., M.R. Meiners, and B.O. Burwell
1992 *A Synthesis and Critique of Studies on Medicaid Asset Spenddown*. January. U.S. Department of Health and Human Services. Available: <http://aspe.hhs.gov/daltcp/reports/syncri.htm> [accessed Nov. 2006].
- Adams, P., M. Hurd, D. McFadden, A. Merrill, and T. Ribeiro
2003 Healthy, wealthy, and wise? Tests for direct causal paths between health and socio-economic status. *Journal of Econometrics* 112(1):3-56.
- Adda, J., T. Chandola, and M. Marmot
2003 Socio-economic status and health: Causality and pathways. *Journal of Econometrics* 112(1):57-63.
- American Sociological Association
2006 Social science consortium hosts U.S. science advisor, NSF, and NIH social science leadership. *Footnotes* 34(1):1, 4.
- Arber, S, J. McKinlay, A. Adams, L. Marceau, C. Link, and A. O'Donnell
2006 Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: A video-simulation experiment. *Social Science and Medicine* 62(1):103-115.
- Arkes, H.R.
2003 The nonuse of psychological research at two federal agencies. *Psychological Science* 14:1-6.
- Arrow, K.
1962 Economic welfare and the allocation of resources for invention. Pp. 609-626 in Richard R. Nelson, ed., *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton, NJ: Princeton University Press.

- Ash, M.
2003 Psychology. Pp. 251-274 in T. Porter and D. Ross, eds., *The Cambridge History of Science, Vol. 7. The Modern Social Sciences*. Cambridge, England: Cambridge University Press.
- Association for Psychological Science
2005 NIH advisory panel calls for stable home for basic behavioral science. *Association for Psychological Science Observer* 18(February).
- Balas, E., and S. Boren
2000 Managing clinical knowledge for health care improvement. Pp. 65-70 in V. Schattauer, ed., *Yearbook of Medical Informatics*. Stuttgart, Germany: Schattauer Publishing.
- Bales, R.F.
1950 *Interaction Process Analysis: A Method for the Study of Small Groups*. Cambridge, MA: Addison-Wesley.
- Barker, D.J.P., J.G. Eriksson, T. Forsén, and C. Osmond
2002 Fetal origins of adult disease: Strength of effects and biological basis. *International Journal of Epidemiology* 31:1235-1239.
- Behavioral and Social Research Program, National Institute on Aging
2004 Behavioral and Social Research Program. (Informational CD-Rom, April, Version 6.0). Bethesda, MD: National Institute on Aging.
- Behn, R.D., and J.V. Vaupel
1982 *Quick Analysis for Busy Decision Makers*. New York: Basic Books.
- Ben-David, J.
1971 *The Scientist's Role in Society*. Englewood Cliffs, NJ: Prentice-Hall.
- Berg, J., and T. Rietz
2003 Prediction markets as decision support systems. *Information Systems Frontiers* 5(1):79-93.
- Bernal, J.D.
1939 *The Social Function of Science*. New York: Macmillan.
- Blank, R.M.
1991 The effects of double-blind versus single-blind reviewing: Experimental evidence from the *American Economic Review*. *American Economic Review* 81:1041-1067.
- Boix-Mansilla, V., and H. Gardner
2003 Assessing interdisciplinary work at the frontier. An empirical exploration of "symptoms of quality." Available: <http://www.interdisciplines.org/interdisciplinarity/papers/6/2> [accessed Nov. 2006].
- Boix-Mansilla, V., I. Feller, and H. Gardner
2006 Quality assessment in interdisciplinary research and education. *Research Evaluation* 15(1):69-74(6).
- Bonvillian, W.B., and K.V. Sharp
2001 Homeland security technology. *Issues in Science and Technology* Winter. Available: <http://www.issues.org/18.2/bonvillian.html> [accessed Nov. 2006].
- Boomsma, D., G. Willemsen, C.V. Dolan, L.C. Hawkey, and J.T. Cacioppo
2005 Genetic and environmental contributions to loneliness in adults: The Netherlands Twin Register Study. *Behavior Genetics* 35(6):745-752.
- Bornmann, L., and H.-D. Daniel
2005 Selecting scientific excellence through committee peer review—A citation analysis of previous publications by successful and non-successful post-doctoral research fellow applicants. *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics*, July 24-28, Stockholm, Sweden.

- Bourdieu, P.
1988 *Homo Academicus*. (Translated by Peter Collier). Stanford, CA: Stanford University Press.
- Bowler, P., and I. Morus
2005 *Making Modern Science*. Chicago, IL: University of Chicago Press.
- Boyack, K., and K. Börner
2003 Indicator assisted evaluation of funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology* 54:447-461.
- Bozeman, B., and J. Melkers, eds.
1993 *Evaluating R&D Impacts: Methods and Practice*. Boston, MA: Kluwer Academic Publishers.
- Bozeman, B., and D. Sarewitz
2005 Public values and public failure in U.S. science policy. *Science and Public Policy* 32:119-136.
- Brenneis, D.
1994 Discourse and discipline at the National Research Council: A bureaucratic bildungsroman. *Cultural Anthropology* 9:23-36.
- Bretschneider, S.
1993 Operations research contributions to evaluation of R&D projects. Pp. 123-153 in B. Bozeman and J. Melkers, eds., *Evaluating R&D Impacts: Methods and Practice*. Boston, MA: Kluwer Academic Publishers.
- Bromley, A.
2003 What criteria should be used to establish funding priorities? *Physics Today* 56(6):54-55.
- Brooks, H.
1965 Future needs for the support of basic research. Pp. 77-110 in National Research Council, *Research and National Goals*. Report to the Committee on Science and Astronautics, U.S. House of Representatives. Washington, DC: Superintendent of Documents, U.S. Government Printing Office.
- Brunner, E.J., H. Hemingway, B.R. Walker, M. Page, P. Clarke, M. Juneja, M.J. Shipley, M. Kumari, R. Andrew, J.R. Seckl, A. Papadopoulos, S. Checkley, A. Rumley, G.D. Lowe, S.A. Stansfield, and M.G. Marmot
2002 Adrenocortical, autonomic and inflammatory causes of the metabolic syndrome: Nested case-control study. *Circulation* 106:2659-2665.
- Bruun, H., J. Hukkinen, K. Huutoniemi, and J. Thompson Klein
2005 *Promoting Interdisciplinary Research: The Case of the Academy of Finland*. Helsinki: Academy of Finland.
- Butler, L.
2004 What happens when funding is linked to publication counts? Pp. 389-405 in H. Moed, W. Glanzel, and U. Schmooch, eds., *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Camic, C., and N. Gross
1998 Contemporary developments in sociological theory: Current projects and conditions of possibility. *Annual Review of Sociology* 24:453-476.
- Campanario, J.
1998a Peer review for journals as it stands today: Part 1. *Science Communication* 19:181-211.
1998b Peer review for journals as it stands today: Part 2. *Science Communication* 19:277-306.

- Caplan, N.
1976 Social research and national policy: What gets used, by whom, for what purposes, and with what effects? *International Social Science Journal* 28(1):187-194.
- Carey, J.R.
2003 *Longevity: The Biology and Demography of Life Span*. Princeton, NJ: Princeton University Press.
- Carey J.R., and S. Tuljapurkar, eds.
2003 Life span: Evolutionary, ecological and demographic perspectives. *Population and Development Review* 29 (Suppl):1-320.
- Carey, J.R., P. Liedo, H.G. Muller, J.L. Wang, Y. Zhang, and L. Harshman
2005 Stochastic dietary restriction using a Markov-chain feeding protocol elicits complex, life history response in *Medflies*. *Aging Cell* 4(1):31-39.
- Cargill, S.L., J.R. Carey, H.G. Muller, and G. Anderson
2003 Age of ovary determines remaining life expectancy in old ovariectomized mice. *Aging Cell* 2:185-190.
- Cerf, C., and V. Navasky
1984 *The Experts Speak: The Definitive Compendium of Authoritative Misinformation*. New York: Pantheon Books.
- Cetina, K.K.
1999 *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Chernew, M.E., D.P. Goldman, F. Pan, and B.P. Shang
2005 Disability and health care spending among Medicare beneficiaries. *Health Affairs* 24:W5R42-WR4R52.
- Christensen, K., H. Støvring, and M. McGue
2001a Do genetic factors contribute to the association between birth weight and blood pressure? *Journal of Epidemiology and Community Health* 55:583-587.
- Christensen K., A. Wienke, A. Skytthe, N.V. Holm, J.W. Vaupel, and A.I. Yashin
2001b Cardiovascular mortality in twins and the fetal origins hypothesis. *Twin Research* 4(5):344-349.
- Chubin, D., and E. Hackett
1990 *Peerless Science*. Albany: State University of New York Press.
- Clemen, R.
1991 *Making Hard Decisions: An Introduction to Decision Analysis*. Boston, MA: PWS-Kent.
- Clements, E., W. Powell, K. McIlwaine, and D. Okamoto
1995 Careers in print: Books, journals, and scholarly publications. *American Journal of Sociology* 101:433-494.
- Cohen, I.B.
1985 *Revolution in Science*. Cambridge, MA: Harvard University Press.
- Cole, J., and S. Cole
1981 *Peer Review in the National Science Foundation: Phase Two of a Study*. Washington, DC: National Academy Press.
- Cole, S.
1983 The hierarchy of the sciences? *American Journal of Sociology* 89:111-139.
1992 *Making Science: Between Nature and Society*. Cambridge, MA: Harvard University Press.
2001 Why sociology doesn't make progress like the natural sciences. Pp. 37-60 in S. Cole, ed., *What's Wrong with Sociology*. New Brunswick, NJ: Transaction Publishers.

- Cole, S., L. Rubin, and J. Cole
1978 *Peer Review in the National Science Foundation: Phase One of a Study*. Washington, DC: National Academy Press.
- Cole, S., J. Cole, and G. Simon
1981 Chance and consensus in peer review. *Science* 214:881-886.
- Colwell, R.
1998 The National Science Foundation's Role in the Arctic. Paper presented at Opportunities in Arctic Research: A Community Workshop, September 3, Arlington, VA.
- Committee on Science, Engineering, and Public Policy
1993 *Science, Technology and the Federal Government: National Goals for a New Era*. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Washington, DC: National Academy Press.
- 1997 *International Benchmarking of U.S. Mathematics Research*. Panel on International Benchmarking of U.S. Mathematics Research, National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Washington, DC: National Academy Press.
- 1998 *International Benchmarking of U.S. Materials Science and Engineering Research*. Panel on International Benchmarking of U.S. Mathematics Research. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Washington, DC: National Academy Press.
- 1999a *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Washington, DC: National Academy Press.
- 1999b *International Benchmarking of U.S. Immunology Research*. Panel on International Benchmarking. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Washington, DC: National Academy Press.
- 2000 *Experiments in International Benchmarking of U.S. Research Fields*. Committee on Science, Engineering, and Public Policy. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Washington, DC: National Academy Press.
- 2004 *Setting Priorities for Large Research Facility Projects Supported by the National Science Foundation*. Committee on Setting Priorities for NSF-Sponsored Large Research Facility Projects. Policy and Global Affairs Division. Board on Physics and Astronomy, Division on Engineering and Physical Sciences. Washington, DC: The National Academies Press.
- Committee on Science and Technology, U.S. House of Representatives
1982 *Setting Priorities for Science*. (Hearings). Washington, DC: U.S. Government Printing Office.
- Costa, D.
2003 Understanding mid-life and older age mortality declines: Evidence from Union Army veterans. *Journal of Econometrics* 112:175-192.
- Cozzens, S.
1997 The knowledge pool: Measurement challenges in evaluating fundamental research programs. *Evaluation and Program Planning* 20:77-89.
- Cozzens, S.E., K. Bobb, and I. Bortagaray
2002 Evaluating the distributional consequences of science and technology policies and programs. *Research Evaluation* 11:101-107.
- Crane, D.
1972 *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago, IL: University of Chicago Press.

- Cutler, D.M.
2003 Aging, natural death and the compression of morbidity. *New England Journal of Medicine* 349(11):1048-1055.
- David, P.
1994 Difficulties in assessing the performance of research and development programs. Pp. 293-302 in A. Teich, S. Nelson, and C. McEnaney, eds., *AAAS Science and Technology Policy Yearbook—1994*. Washington, DC: American Association for the Advancement of Science.
- Deaton, A.
2002 Policy implications of the gradient of health and wealth. *Health Affairs* 21:13-30.
- Deutsch, K.W., J. Platt, and D. Senghaas
1971 Conditions favoring major advances in social science. *Science* 171:450-459.
- Dirk, L.
1999 A measure of originality: The elements of science. *Social Studies of Science* 29:765-776.
- Dogan, M., and R. Pahre
1990 *Creative Marginality: Innovation at the Intersections of Social Sciences*. Boulder, CO: Westview.
- Edwards, W.
1954 The theory of decision making. *Psychological Bulletin* 51:380-417.
- Espeland, W.N., and M.L. Stevens
1998 Commensuration as a social process. *Annual Review of Sociology* 24:313-343.
- Evenson, R., P. Waggoner, and V. Ruttan
1979 Economic benefits from research: An example from agriculture. *Science* 205:1101-1107.
- Ezrahi, Y.
1978 Political contexts of science indicators. Pp. 285-322 in Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, and H. Zuckerman, eds., *Toward A Metric of Science, The Advent of Science Indicators*. New York: Wiley.
- Ezzati, M., A.D. Lopez, A. Rodgers, S. Vanderhorn, C.J.L. Murray, and the Comparative Risk Assessment Collaborating Group
2002 Selected major risk factors and global and regional burden of disease. *Lancet* 360:1347-1360. Available: <http://lib.haifa.ac.il/www/mesila/nutrition/gurmi%20sikun.pdf> [accessed Nov. 2006].
- Feller, I.
1992 Alternative Models of Research Performance. Final Report, Office of Technology Assessment, Contract Number I3-4065.0.
1996 The determinants of research competitiveness among universities. Pp. 35-72 in A. Teich, ed., *Competitiveness in Academic Research*. Washington, DC: American Association for the Advancement of Science.
2002 Performance measurement redux. *American Journal of Evaluation* 23:435-452.
2006 Multiple actors, multiple settings, multiple criteria: Issues in assessing interdisciplinary research. *Research Evaluation* 15(1):5-15.
- Fischhoff, B.
2000 Scientific management of science? *Policy Sciences* 33:73-87.
- Fisher, E.S., D.E. Weinberg, T.A. Stukel, D.J. Gottlieb, F.L. Lucas, and E.L. Pinder
2003a The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of health care. *Annals of Internal Medicine* 138:273-287.
2003b The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine* 138:288-298.

- Freedman, V.A., H. Aykan, and L.G. Martin
2001 Aggregate changes in severe cognitive impairment among older Americans: 1993 and 1998. *Journal of Gerontology B: Social Sciences* 56(2):S100-S111.
- Freedman, V.A., L.G. Martin, and R.F. Schoeni
2002 Recent trends in disability and functioning among older adults in the United States: A systematic review. *Journal of the American Medical Association* 288:3137-3146.
- Freeman, R., E. Weinstein, E. Marincola, J. Rosenbaum, and F. Solomon
2001 Competition and careers in biosciences. *Science* 294:2293-2294.
- Fri, R.
2004 *Methodology for Estimating Prospective Benefits of Energy Efficiency and Fossil Energy R&D*. Letter to D. Garman and M. Maddox, U.S. Department of Energy on behalf of National Research Council Board on Energy and Environmental Systems, November 15.
- Frist, W.
2002 Federal funding for biomedical research: Commitment and benefits. *Journal of the American Medical Association* 287:1722-1724.
- Furner, M.
1975 *Advocacy and Objectivity*. Lexington: University Press of Kentucky.
- Galison, P.
1999 Trading zone: Coordinating action and beliefs. Pp. 137-160 in M. Baglioli, ed., *The Science Studies Reader*. London, England: Routledge.
- Garfinkel, H.
1967 *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Geisler, E.
2000 *The Metrics of Science and Technology*. Westport, CT: Quorum Books.
- Gerstein, D.
1986 Introduction. Pp. 1-20 in National Research Council, *Behavioral and Social Sciences, Fifty Years of Discovery*. N. Smelser and D. Gerstein, eds. Washington, DC: National Academy Press.
- Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow
1994 *The New Production of Knowledge*. London, England: Sage Publications.
- Glanzel, W., and A. Schubert
2004 Analyzing scientific networks through co-authorship. Pp. 257-276 in H. Moed, W. Glanzel, and U. Schmooch, eds., *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Glei, D.A., D.A. Landau, N. Goldman, Y.L. Chuang, G. Rodriguez, and M. Weinstein
2005 Participating in social activities helps preserve cognitive function: An analysis of a longitudinal, population-based study of the elderly. *International Journal of Epidemiology* 34:864-871.
- Godin, B.
2002 Outline for a history of science measurement. *Science, Technology and Human Values* 27:3-27.
- Goldman, D.P., B.P. Shang, J. Bhattacharya, A.M. Garber, M. Hurd, G.F. Joyce, D.N. Lakdawalla, C. Panis, and P.G. Shekele
2005 Consequences of Health Trends and Medical Innovation for the Future Elderly. Available: <http://content.healthaffairs.org/cgi/content/abstract/hlthaff.w5.r5> [accessed Nov. 2006].
- Gramlich, E.
1981 *Benefit Cost Analysis of Government Programs*. Englewood Cliffs, NJ: Prentice-Hall.

- Greenberg, D.
2001 *Science, Money, and Politics: Political Triumph and Ethical Erosion*. Chicago, IL: University of Chicago Press.
- Gross, C.P., G.F. Anderson, and N.R. Powe
1999 The relation between funding by the National Institutes of Health and the burden of disease. *New England Journal of Medicine* 340(24):1914-1915.
- Grupp, H., and M. Moge
2004 Indicators for national science and technology policy. Pp. 75-94 in H. Moed, W. Glanzel, and U. Schmooch, eds., *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Guetzkow, J., M. Lamont, and G. Mallard
2004 What is originality in the social sciences and the humanities? *American Sociological Review* 69:190-212.
- Guston, D.
2000 *Between Politics and Science: Assuring the Integrity and Productivity of Research*. Cambridge, England: Cambridge University Press.
- Guston, D., and K. Keniston, eds.
1994 *The Fragile Contract: University Science and the Federal Government*. Cambridge, MA: MIT Press.
- Hagstrom, W.
1965 *The Scientific Community*. New York: Basic Books.
- Hall, B.
1996 The private and social rates of returns to research and development. Pp. 140-162 in B. Smith and C. Barfield, eds., *Technology, R&D, and the Economy*. Washington, DC: Brookings Institution/American Enterprise Institute.
- Hargens, L.
1987 Scholarly consensus and journal selection rates. *American Sociological Review* 53:174-196.
- Hart, R.
1999 On the problem of Chinese science. Pp. 189-201 in M. Biagioli, ed., *The Science Studies Reader*. New York: Routledge.
- Hatry, H.
1989 Determining the effectiveness of government services. Pp. 469-482 in J. Perry, ed., *Handbook of Public Administration*. San Francisco, CA: Jossey-Bass.
- Hawley, L.C., C.M. Masi, J.D. Berry, and J.T. Cacioppo
2006 Loneliness is a unique predictor of age-related differences in systolic blood pressure. *Psychology and Aging* 21:152-164.
- Henkel, M.
1999 The modernization of research evaluation: The case of the UK. *Higher Education* 38:105-122.
- Hicks, D.
2004 The four literatures of social science. Pp. 473-495 in H. Moed, W. Glanzel, and U. Schmooch, eds., *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Hicks, D., H. Tomizawa, Y. Saitoh, and S. Kobayashi
2004 Bibliometric evaluation of federally funded research in the United States. *Research Evaluation* 13(2):76-86.
- Hollingsworth, J.R.
2003 Research Organizations and Major Discoveries in Twentieth-Century Science: A Case Study of Excellence in Biomedical Research. (Paper P-02-003). Available: <http://skylla.wz-berlin.de/pdf/2002/p02-003.pdf> [accessed Nov. 2006].

- Hollingsworth, J.R., and E.J. Hollingsworth
2000 Major discoveries and biomedical research organizations: Perspectives on interdisciplinarity, nurturing leadership, and integrated structure and cultures. Pp. 215-244 in P. Weingart and N. Stehr, eds., *Practising Interdisciplinarity*. Toronto: University of Toronto Press.
- Hook, E., ed.
2002 *Prematurity in Scientific Discovery*. Berkeley: University of California Press.
- Horan, J.
1996 *The End of Science*. Reading, MA: Addison-Wesley.
- Howard, R.A., and J.E. Matheson, eds.
1989 *The Principles and Applications of Decision Analysis*. Menlo Park, CA: Strategic Decisions Group.
- Hull, D.L.
1988 *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago, IL: University of Chicago Press.
- Illinois Institute of Technology
1968 *Technology in Retrospective and Critical Events*. (A research report prepared for the National Science Foundation under Contract NSF-C535). Chicago: Illinois Institute of Technology.
- Industrial Research Institute
1999 *Measuring and Improving the Performance and Return on R&D*. Washington, DC: Industrial Research Institute.
- Inkeles, A.
1986 Advances in sociology: A critique. Pp. 13-31 in K.W. Deutsch, A.S. Markovits, and J. Platt., eds., *Advances in the Social Sciences: What, Who, Where, How?* Lanham, MD: University Press of America.
- Institute of Medicine
1991 *Extending Life, Enhancing Life: A National Research Agenda on Aging*. Committee on a National Research Agenda on Aging. E.T. Lonergan, ed. Washington, DC: National Academy Press.
1998 *Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at the National Institutes of Health*. Committee on the NIH Research Priority-Setting Process. Washington, DC: National Academy Press.
2004 *NIH Extramural Center Programs: Criteria for Initiation and Evaluation*. Committee for Assessment of NIH Centers of Excellence Programs. Board on Health Sciences Policy. F.J. Manning, M. McGeary, and R. Estabrook, eds. Washington, DC: The National Academies Press.
- Irvine, J., and B.R. Martin
1984 *Foresight in Science: Picking the Winners*. London, England: Frances Pinter.
- Jaeger, C.C., O. Renn, E.A. Rosa, and T. Webler
2001 *Risk, Uncertainty and Rational Action*. London, England: Earthscan.
- Jaffe, A., and M. Trajtenberg
2002 *Patents, Citations, and Innovations*. Cambridge, MA: MIT Press.
- Jaschik, S
2006 Anti-gay “flagging” at NEH? *Inside Higher Ed*, January 9. Available: <http://insidehighered.com/news/2006/01/09/neh> [accessed Nov. 2006].
- Jones, E.
1988 *Growth Recurring*. Oxford, England: Oxford University Press.
- Juster, T.F., and R. Suzman
1995 An overview of the Health and Retirement Study. *Journal of Human Resources (Suppl)*:S7-S56.

- Keeney, R.L., and H. Raiffa
1976 *Decision with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley and Sons.
- Khazaeli, A.A., and J.W. Curtsinger
2001 Genetic analysis of extended lifespan in *Drosophila melanogaster*. III. On the relationship between artificially selected lines and wild stocks. *Genetica* 109:245-253.
- Klein, J.T.
1996 *Crossing Boundaries: Knowledge, Disciplinarity, and Interdisciplinarity*. Charlottesville: University Press of Virginia.
- Kohler, R.E.
1987 Science, foundations, and American universities in the 1920s. *Osiris* 3:135-164.
1991 *Partners in Science: Foundations and Natural Scientists, 1900-1945*. Chicago, IL: University of Chicago Press.
2002 *Landscapes and Labyrinths*. Chicago, IL: University of Chicago Press.
- Kreilkamp, K.
1971 Hindsight and the real world of science policy. *Science Studies* 1:43-66.
- Krige, J.
1980 *Science, Revolution and Discontinuity*. Brighton, England: Harvester Press.
- Kuhn, T.S.
1962 *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Laband, D., and M. Pietter
1994 The relative impacts of economic journals: 1970-1990. *Journal of Economic Literature* 32:640-666.
- Lakdawalla, D.N., D.P. Goldman, J. Bhattacharya, M. Hurd, G. Joyce, and C. Panis
2003 Forecasting the nursing home population. *Medical Care* 41(1):8-20.
- Lakdawalla, D.N., D.P. Goldman, and B.P. Shang
2005 The health and cost consequences of obesity among the future elderly. *Health Affairs* 24(Suppl):W5R30-W5R41.
- Lamont, M.
2004 Theoretical growth and conceptual foreplay. *Perspectives, Newsletter of the ASA Theory Section* 27(3):8.
- Lamont, M., and G. Mallard
2005 *Peer Evaluation in the Social Sciences and the Humanities Compared: The United States, the United Kingdom, and France*. Report prepared for the Social Sciences and Humanities Research Council of Canada, March 24.
- Lamont, M., G. Mallard, and J. Guetzkow
2006 Beyond blind faith: Overcoming the obstacles to interdisciplinary evaluation. *Research Evaluation* 15(1):43-55.
- Landry, R., M. Lamari, and N. Amara
2003 The extent and determinants of the utilization of university research in government agencies. *Public Administration Review* 63(2):192-205.
- Langfeldt, L.
2001 The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science* 31:820-841.
- Latour, B., and S. Woolgar
1979 *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills, CA: Sage Publications.
- Laudel, G.
2006 Conclave in the tower of Babel: How peers review interdisciplinary research proposals. *Research Evaluation* 15(1):57-68.

- Lee, R.D.
2003 Rethinking evolutionary theory of aging: Transfers, not births, shape senescence in social species. *Proceedings of the National Academy of Sciences* 100:9637-9642.
- Levy, B.R.
2003 Mind matters: Cognitive and physical effects of aging self-stereotypes. *Journal of Gerontology: Psychological Sciences* 58B(4):P203-P211.
- Lindblom, C.E., and D.K. Cohen
1979 *Usable Knowledge: Social Science and Social Problem Solving*. New Haven, CT: Yale University Press.
- Link, A.
1996 *Evaluating Public Sector Research and Development*. New York: Praeger.
- Losee, R.
1995 The development and migration of concepts from donor to borrower disciplines: Sublanguage term use in hard and soft sciences. Pp. 265-274 in *Proceedings of the Fifth International Conference on Scientometrics and Informetrics*, June, Chicago, IL.
- Loucks, E.B., L.F. Berkman, T.L. Gruenewald, and T.E. Seeman
2005 Social integration is associated with fibrinogen concentration in elderly men. *Psychosomatic Medicine* 67:353-358.
- Lubitz, J., L. Cai, E. Kramarow, and H. Lentzner
2003 Health, life expectancy, and health care spending among the elderly. *New England Journal of Medicine* 349:1048-1055.
- Lynn, L.E., ed.
1978 *Knowledge and Policy: The Uncertain Connection*. Washington, DC: National Academy Press.
- MacRae, D., Jr., and I. Feller
1998 The structure of and prospects for policy research as suggested by journal citation analysis. *Policy Studies Review* 15:115-135.
- Mallard, G., M. Lamont, and J. Guetzkow
2005 Epistemological pluralism and fairness in peer evaluation: Evidence from the social sciences and the humanities. Submitted to *American Journal of Sociology*.
- Mandel, H.G., and E.S. Vesell
2006 Declines in funding of NIH R01 research grants. *Science* 313:1387-1388.
- Mansilla, V.B., and H. Gardner
2004 Assessing interdisciplinary work at the frontier: An empirical exploration of "symptoms of quality." December 28. Available: <http://www.interdisciplines.org/interdisciplinarity/papers/6/printable/paper> [accessed Nov. 2006].
- Manton, K.G., and X.L. Gu
2001 Changes in the prevalence of chronic disability in the United States black and non-black populations above age 65 from 1982 to 1999. *Proceedings of the National Academy of Sciences* 98:6354-6359.
- Manton, K.G., and A.I. Yashin
2000 *Mechanisms of Aging and Mortality: Searches for New Paradigms*. (Monographs on Population Aging 7). Odense, Denmark: Odense University Press.
- Marburger, J.
2005 Address to the 30th Annual AAAS Forum on Science and Technology Policy, April 21, Washington, DC. Available: <http://www.aaas.org/news/releases/2005/0421marburgerText.shtml> [accessed Nov. 2006].
- Martin, B.
1996 Technology foresight: Capturing the benefits from science-related technologies. *Research Evaluation* 6(2):158-168.

- McKenna, M.T., C.M. Michaud, C.J.L. Murray, and J.S. Marks
2005 Assessing the burden of disease in the United States using disability-adjusted life-years. *American Journal of Preventive Medicine* 28(5):415-423.
- Merton, R.
1965 *On the Shoulders of Giants*. Chicago, IL: University of Chicago Press.
- Moed, H., W. Glanzel, and U. Schmooch, eds.
2004 *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Monastersky, R.
2005 The number that's devouring science. *Chronicle of Higher Education*, October 14. Available: <http://chronicle.com/free/v52/i08/08a01201.htm> [accessed Nov. 2006].
- Morgan, M.G., and M. Henrion
1990 *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, England: Cambridge University Press.
- Morillo, F., M. Bordon, and I. Gomez
2001 An approach to interdisciplinarity through bibliometric indicators. *Scientometrics* 51:203-222.
- Mroczek, D., and A. Spiro
2003 Modelling intraindividual change in personality traits: Findings from the normative aging study. *Journal of Gerontology: Psychological Sciences* 58B(3):P153-P165.
- Murphy, K., and R. Topel
1999 The economic value of medical research. Pp. 41-73 in K. Murphy and R. Topel, eds., *Measuring the Gains from Medical Research. An Economic Approach*. Chicago, IL: University of Chicago Press.
- Narin, F., K. Hamilton, and D. Olivastro
1997 The increasing linkage between U.S. technology and public service. *Research Policy* 26(3):317-330.
- Nasar, S.
1998 *A Beautiful Mind*. New York: Simon and Schuster.
- National Academy of Public Administration
1999 *Remembering the Future: Applying Foresight Techniques to Research Planning at EPA*. Washington, DC: National Academy of Public Administration.
- National Academy of Sciences
1965 Basic research and national goals. A report to the U.S. Congress, House of Representatives, Committee on Science and Astronautics. *Minerva* (3)4:499.
- National Institute on Aging
2001 *Action Plan for Aging Research: Strategic Plan for Fiscal Years 2001-2005*. Bethesda, MD: National Institute on Aging. Available: <http://www.nia.nih.gov/NR/rdonlyres/E6765778-B533-44BB-9774-1FB6821B1A14/2696/niasp.pdf> [accessed Nov. 2006].
2005 *Annual Data Report: Fiscal Year 2005*. Available: <http://www.nia.nih.gov/About-NIA/NACA/MeetingInformation> [accessed Nov. 2006].
- National Research Council
1972 *Physics in Perspective: The Nature of Physics and the Subfields of Physics*. Physics Survey Committee. Washington, DC: National Academy Press.
1986 *Behavioral and Social Sciences, Fifty Years of Discovery*. Committee on Basic Research in the Behavioral and Social Sciences. N.J. Smelser and D.R. Gerstein, eds. Washington, DC: National Academy Press.
1988 *The Behavioral and Social Sciences: Achievements and Opportunities*. Committee on Basic Research in the Behavioral and Social Sciences. D.R. Gerstein, R.D. Luce, N.J. Smelser, and S. Sperlrich, eds. Washington, DC: National Academy Press.

- 1995a *Allocating Federal Funds for Science and Technology*. Committee on Criteria for Federal Support of Research and Development. Washington, DC: National Academy Press.
- 1995b *Research Doctorate Programs in the United States: Continuity and Change*. Committee for the Study of Research–Doctorate Programs in the United States. M.L. Goldberger, B.A. Maher, and P. Ebert Flattau, eds. Washington, DC: National Academy Press.
- 1996 *Understanding Risk: Informing Decisions in a Democratic Society*. Committee on Risk Characterization. P.C. Stern and H.V. Fineberg, eds. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- 1997 *Between Zeus and the Salmon: The Biodemography of Longevity*. K.W. Wachter and C.E. Finch, eds. Committee on Population, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- 1998 *Trends in the Early Careers of Life Scientists*. Committee on Dimensions, Causes, and Implications of Recent Trends in the Careers of Life Scientists. Board on Biology, Commission on Life Sciences, Office of Scientific and Engineering Personnel. Washington, DC: National Academy Press.
- 1999 *Making Climate Forecasts Matter*. P.C. Stern and W.E. Easterling, eds. Panel on the Human Dimensions of Seasonal-to-Interannual Climate Variability, Committee on the Human Dimensions of Global Change. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- 2000a *The Aging Mind: Opportunities in Cognitive Research*. Committee on Future Directions for Cognitive Research on Aging. P.C. Stern and L.L. Carstensen, eds. Board on Behavioral, Cognitive, and Sensory Sciences Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- 2000b *Bridging Disciplines in the Brain, Behavioral, and Clinical Sciences*. T.C. Pellmar and L. Eisenberg, eds. Committee on Building Bridges in the Brain, Behavioral, and Clinical Sciences. Division of Neuroscience and Behavioral Health. Washington, DC: National Academy Press.
- 2001a *Energy Research at DOE: Was it Worth it? Energy Efficiency and Fossil Energy Research, 1978 to 2000*. Committee on Benefits of DOE R&D on Energy Efficiency and Fossil Energy, Board on Energy and Environmental Systems. Division on Engineering and Physical Sciences. Washington, DC: National Academy Press.
- 2001b *Grand Challenges in Environmental Sciences*. Committee on Grand Challenges in Environmental Sciences. Washington, DC: National Academy Press.
- 2001c *New Horizons in Health: An Integrative Approach*. Committee on Future Directions for Behavioral and Social Sciences Research at the National Institutes of Health. B.H. Singer and C.D. Ryff, eds. Board on Behavioral, Cognitive, and Sensory Sciences, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- 2001d *Cells and Surveys: Should Biological Measures Be Included in Social Science Research?* Committee on Population, Commission on Behavioral and Social Sciences and Education. C.E. Finch, J.W. Vaupel, and K. Kinsella, eds. Washington, DC: National Academy Press.
- 2003 *Assessing Research-Doctorate Programs: A Methodology Study*. J.P. Ostriker and C.V. Kuh, eds. Assisted by J.A. Voytuk. Committee to Examine the Methodology for the Assessment of Research-Doctorate Programs. Policy and Global Affairs Division Washington, DC: The National Academies Press.
- 2005a *Decision Making for the Environment: Social and Behavioral Science Research Priorities*. Panel on Social and Behavioral Science Priorities for Environmental Decision Making, Committee on the Human Dimensions of Global Change. G.D. Brewer and P.C. Stern, eds. Washington, DC: The National Academies Press.

- 2005b *Prospective Evaluation of Applied Energy Research and Development at DOE (Phase One): A First Look Forward*. Committee on Prospective Benefits of DOE's Energy Efficiency and Fossil Energy R&D Programs. Board on Energy and Environmental Systems, Division on Engineering and Physical Sciences. Washington, DC: The National Academies Press.
- 2005c *Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program*. Committee on Metrics for Global Change Research. Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies. Washington, DC: The National Academies Press.
- National Science Board
- 2004 *Science and Engineering Indicators*. Arlington, VA: National Science Foundation.
- National Science Foundation Advisory Committee for GPRA Performance Assessment
- 2004 *Report of the Advisory Committee for GPRA Performance Assessment*. Available: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf04216 [accessed Jan. 2007].
- National Science and Technology Council
- 1996 *Assessing Fundamental Science*. Washington, DC: Office of the President.
- Nelson, R.
- 1959 The simple economics of basic scientific research. *Journal of Political Economy* 67(2):297-306.
- Nelson, R., ed.
- 1993 *National Innovation Systems*. New York: Oxford University Press.
- North, D.W., and O. Renn
- 2005 *Decision Analytic Tools and Participatory Decision Processes*. Paper prepared for the National Research Council Panel on Public Participation in Environmental Assessment and Decision Making Workshop and Panel Meeting #3, February 3-5, Washington, DC.
- Nye, M. J.
- 1993 *From Chemical Philosophy to Theoretical Chemistry*. Berkeley: University of California Press.
- Oh, C.H.
- 1996 *Linking Social Science Information to Policy Making*. Greenwich, CT: JAI Press.
- Olsen, R., and D. Levy
- 2004 Program Performance and the President's Budget: Do OMB's PART Scores Really Matter? Paper presented at the Association for Policy Analysis and Management's 2004 Annual Research Conference.
- Olshansky, S.J., D.J. Passaro, R.C. Hershow, J. Layden, B.A. Carnes, J. Brody, L. Hayflick, R.N. Butler, D.B. Allison, and D.S. Ludwig
- 2005 A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine* 352(11):1138-1145.
- Parsons, T.
- 1937 *The Structure of Social Action: A Study in Social Theory with Special Reference to a Group of Recent European Writers*. New York: McGraw-Hill.
- 1964 *The Social System*. New York: Free Press.
- Perrin, B.
- 1998 Effective use and misuse of performance measurement. *American Journal of Evaluation* 19:367-379.
- Pichler, R.
- 2006 The ERC and policy maker's expectations: Evaluation as a change agent. *Platform Special Edition*, No. 28(April):3-10.

- Pinkau, K., and O. Renn, eds.
1998 *Environmental Standards: Scientific Foundations and Rational Procedures of Regulation with Emphasis on Radiological Risk Management*. Boston, MA: Kluwer Academic Publishers.
- Pletcher, S.D., and J.W. Curtsinger
2000a The influence of environmentally induced heterogeneity on age-specific genetic variance for mortality rates. *Genetic Research* 75:321-329.
- Pletcher, S.D., and J.W. Curtsinger
2000b Why do lifespans differ? Partitioning mean longevity differences in terms of age-specific mortality parameters. *Journal of Gerontology* 55:B381-389.
- Policy Research in Engineering, Science and Technology
2006 *Evaluation of the United Kingdom Foresight Programme*. Manchester, England: Manchester Business School, University of Manchester.
- Porter, A., and F. Rossini
1985 Peer review of interdisciplinary research. *Science, Technology and Human Values* 10:33-38.
- Power, M.
1997 *The Audit Society*. New York: Oxford University Press.
- Preston, S.H.
2005 Deadweight? The influence of obesity on longevity (editorial). *New England Journal of Medicine* 352(11):1135-1137.
- Price, D.J. De Solla
1963 *Little Science, Big Science*. New York: Columbia University Press.
- Radin, B.
2000 The Government Performance and Results Act and the tradition of federal management reform: Square pegs in round holes? *Journal of Public Administration Research and Theory* 10:111-135.
- Raiffa, H.
1968 *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison-Wesley.
- Rea, S.L., D. Wu, J.R. Cypser, J.W. Vaupel, and T.E. Johnson
2005 A stress-sensitive reporter predicts longevity in isogenic populations of *Caenorhabditis elegans*. *Nature Genetics* 37:894-898.
- Renn, O., T. Webler, and P. Wiedemann
1996 *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Rettig, R., J. Sorig, and B. Milward
1974 Criteria for the Allocation of Resources to Research and Development: A Review of the Literature. Report to the National Science Foundation.
- Reuter, P., and J. Smith-Ready
2002 Editors' note: Assessing JPAM after 20 Years. *Journal of Policy Analysis and Management* 21:339-353.
- Reynolds, S.L., Y. Saito, and E.L. Crimmins
2005 The impact of obesity on active life expectancy in older American men and women. *The Gerontologist* 45(4):438-444.
- Rhoades, G.
2002 Who's doing it right? Strategic activity in public research universities. *Review of Higher Education* 24:41-66.

- Rhoten, D.
2003 *Multi-Method Analysis of the Social and Technical Conditions for Interdisciplinary Collaboration*. Final report prepared for the National Science Foundation (NSF BCS-0129573). Available: http://www.hybridvigor.net/interdis/pubs/hv_pub_interdis-2003.09.29.pdf [accessed Nov. 2006].
- Rinia, E.J., T.N. van Leeuwen, H.G. van Vuren, and A.F.J. van Raaij
2001 Influence on interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy* 30:357-361.
- Roco, M., and W. Bainbridge, eds.
2003 *Nanotechnology: Societal Implications—Maximizing Benefits for Humanity*. Report of the National Nanotechnology Initiative Workshop, December 2-3, 2003. Arlington, VA: National Science Foundation.
- Roessner, D.
2000 Quantitative and qualitative methods and measures in the evaluation of research. *Research Evaluation* 9:125-132.
- Romsdahl, R.J.
2005 When do environmental decision makers use social science? Pp. 139-174 in National Research Council, *Decision Making for the Environment: Social and Behavioral Science Research Priorities*. Panel on Social and Behavioral Science Research Priorities for Environmental Decision Making, G.D. Brewer and P.C. Stern, eds. Washington, DC: The National Academies Press.
- Rooks, N.M.
2006 *White Money/Black Power: The Surprising History of African American Studies and the Crisis of Race in Higher Education*. Boston, MA: Beacon Press.
- Rosekranz, M.A., D.C. Jackson, K.M. Dalton, I. Dolsky, C.D. Ryff, B.S. Singer, D. Muller, N.H. Kalin, and R.J. Davidson
2003 Affective style and in vivo immune response: Neurobehavioral mechanisms. *Proceedings of the National Academy of Sciences* 100:11148-11152.
- Ross, D.
2003 Changing contours of the social science disciplines. Pp. 205-237 in T. Porter and D. Ross, eds., *The Cambridge History of Science, Vol. 7. The Modern Social Sciences*. Cambridge, England: Cambridge University Press.
- Ruegg, R., and I. Feller
2003 *A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade*. (NIST GCR 03-857). Washington, DC: National Institute of Science and Technology.
- Rule, J.B.
1997 *Theory and Progress in Social Science*. Cambridge, England: Cambridge University Press.
- Ruttan, V.
1982 *Agricultural Research Policy*. Minneapolis: University of Minnesota Press.
- Scarpa, T.
2006 Peer review at NIH. *Science* 311:41.
- Scherer, F.
1965 Government research and development programs. Pp. 12-57 in R. Dorfman, ed., *Measuring Benefits of Government Investments*. Washington, DC: Brookings Institution.
- Seidman, H.
1998 *Politics, Position, and Power, 5th ed.* Oxford, England: Oxford University Press.
- Shapin, S.
1996 *The Scientific Revolution*. Chicago: University of Chicago Press.

- Sherwin, C.W., and R. Isenson
1966 First Interim Report on Project Hindsight, Summary, 30 June 1966; revised 13 Oct. 1966, No. AD-642-4000, Clearinghouse for Federal Scientific and Technical Information, Springfield, VA.
- Smelser, N.J.
1986 The Ogburn vision fifty years later. Pp. 21-34 in National Research Council, *Behavioral and Social Sciences, Fifty Years of Discovery*. Committee on Basic Research in the Behavioral and Social Sciences. N.J. Smelser and D.R. Gerstein, eds. Washington, DC: National Academy Press.
2005 The questionable logic of “mistakes” in the dynamics of knowledge growth in the social sciences. *Social Research* 72:237-262.
- Smith, B.
1982 The concept of scientific choice: A brief review of the literature. In *Setting Priorities for Science*, Hearings before the Subcommittee on Science, Research and Technology, Committee on Science and Technology, U.S. House of Representatives, Ninety-seventh Congress, Second Session, September 30; December 8, 1982, Appendix I.
1990 *American Science Policy Since World War II*. Washington, DC: Brookings Institution.
- Steinmetz, G., ed.
2005 *The Politics of Method in the Human Sciences: Positivism and Its Epistemological Others*. Durham, NC: Duke University Press.
- Steinsaltz, D., S.N. Evans, and K.W. Wachter
2005 A generalized model of mutation-selection balance with applications to aging. *Advances in Applied Mathematics* 35:16-33.
- Stent, G.
2002 Prematurity and uniqueness in scientific discovery. Pp. 22-33 in E. Hook, ed., *Prematurity in Scientific Discovery*. Berkeley: University of California Press.
- Stephan, P., and S. Levin
1992 *Striking the Mother Lode in Science*. New York: Oxford University Press.
- Stigler, S.
1993 Competition and the research universities. *Daedalus* 122(4):157-177.
1999 Statistical concepts in psychology. Pp. 189-199 in S. Stigler, ed., *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press.
- Stiglitz, J.E.
1988 *Economics of the Public Sector*, Second Edition. New York: W.W. Norton and Company.
- Stokes, D.
1997 *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution.
- Suzman, R.
2004 Research on population aging at NIA: Retrospect and prospect. Pp. 239-264 in *Ageing, Health, and Public Policy: Demographic and Economic Perspectives*. (Population and Development Review, a supplement to vol. 30). New York: Population Council.
- Tan, Q., G. De Benedictis, A.I. Yashin, L. Bathum, L. Christiansen, J. Dahlggaard, N. Frizner, W. Vach, J.W. Vaupel, K. Christensen, and T.A. Kruse
2004 Assessing genetic association with human survival at multi-allelic loci. *Biogerontology* 4:89-97.

- Thackray, A.
1978 Measurement in the historiography of science. Pp. 11-30 in Y. Elkana, J. Lederberg, R. Merton, A. Thackray, and H. Zuckerman, eds., *Toward a Metric of Science*. New York: John Wiley and Sons.
- Thomas, J.
1999 The Unpredictability of Science and Its Consequences. Presentation to the Royal Institution of Great Britain.
- Toulmin, S.
1964 The complexity of scientific choice: A stocktaking. *Minerva* II (Spring, 3): 343-359.
- Travis, G.D.L., and H.M. Collins
1991 New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, and Human Values* 16:322-341.
- U.S. Congress Office of Technology Assessment
1986 *Research Funding as an Investment: Can We Measure the Returns?* (OTA-TM-SET-36). Washington, DC: U.S. Government Printing Office.
1991 *Federally Funded Research: Decisions for a Decade*. (OTA-SET-490). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Energy Office of Science
2004 *Office of Science Strategic Plan, February 2004*. Available: http://www.er.doe.gov/Sub/Mission/Strategic_Plan/Feb-2004-Strat-Plan-screen-res.pdf [accessed Dec. 2006].
- Van Asselt, M.B.A.
2000 *Perspectives on Uncertainty and Risk: The Prima Approach to Decision Support*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- van Leeuwen, T.
2004 Descriptive versus evaluative bibliometrics. Pp. 375-388 in H. Moed, W. Glanzel, and U. Schmooch, eds., *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- van Leeuwen, T., and R. Tijssen
2000 Interdisciplinary dynamics of modern science: Analysis of cross-disciplinary citation flows. *Research Evaluation* 9:183-186.
- van Raan, A.F.J., ed.
1988a *Handbook of Quantitative Studies of Science and Technology*. Amsterdam, Netherlands: North-Holland.
- van Raan, A.F.J.
1988b Introduction to the handbook. Pp. 1-xx in A.F. J. van Raan, ed., *Handbook of Quantitative Studies of Science and Technology*. Amsterdam, Netherlands: North-Holland.
2004a Measuring science. Pp. 19-50 In H. Moed, W. Glanzel, and U. Schmoch, eds., *Handbook of Quantitative Science and Technology Research*. Dordrecht, Netherlands: Kluwer Academic Publishers.
2004b Sleeping beauties in science. *Scientometrics* 59:461-466.
2005 Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics* 62:133-143.
- Vaupel J.W., J.R. Carey, K. Christensen, T.E. Johnson, A.I. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, P. Liedo, V.D. Longo, Y. Zeng, K.G. Manton, and J.W. Curtsinger
1998 Biodemographic trajectories of longevity. *Science* 280:855-860.
- Vaupel, J.W., J.R. Carey, and K. Christensen
2003 It's never too late. *Science* 301:1679-1681.

- Vogel, G.
2004 European science: New commissioner calls for evolution, not revolution. *Science* 306:1459.
- von Winterfeldt, D., and W. Edwards
1986 *Decision Analysis and Behavioral Research*. Cambridge, England: Cambridge University Press.
- Wagner, C., and L. Leydesdorff
2005 Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34:1608-1618.
- Waite, L.
2004 The demographic faces of the elderly. *Population and Development Review* 30 (Suppl):1-16.
- Weinberg, A.
1963 Criteria for scientific choice. *Minerva* I(2):159-171.
1964 Criteria for scientific choice II: The two cultures. *Minerva* III(1):3-14.
1989 Criteria for evaluation, a generation later. Pp. 3-12 in D. Evered and S. Harnett, eds., *The Evaluation of Scientific Research*. Chichester, England: John Wiley & Sons.
- Weinberg, S.
2001 The non-revolution of Thomas Kuhn. Pp. 187-206 in S. Weinberg, ed., *Facing Up: Science and Its Cultural Adversaries*. Cambridge, MA: Harvard University Press.
- Weingart, P.
2005 Impact of bibliometrics upon the science system: Inadvertent consequences. *Scientometrics* 62:117-131.
- Weiss, C.H.
1977 Research for policy's sake: The enlightenment function of social research. *Policy Analysis* 3(4):531-545.
1979 The many meanings of research utilization. *Public Administration Review* 39(5):426-431.
- Wen, M., K.A. Cagney, and N.A. Christakis
2005 Effect of specific aspects of community social environment on the mortality of individuals diagnosed with serious illness. *Social Science and Medicine* 61:1119-1134.
- Wessely, S.
1996 What do we know about peer review? *Psychological Medicine* 26:883-886.
- Wilmoth, J.R., L.J. Deegin, H. Lundström, and S. Horiuchi
2000 Increase of maximum life-span in Sweden, 1861-1999. *Science* 289:2366-2368.
- Wolfe, A.
2005 Reality in political science. *Chronicle of Higher Education* (November 4):B19.
- Wong, M.D., M.F. Schapiro, J. Boscardin, and S.U. Ettner
2002 Contributions of major diseases to disparities in mortality. *New England Journal of Medicine* 347:1585-1592.
- Zeng, Y., and J.W. Vaupel
2003 Association of late childbearing with healthy longevity among the oldest-old in China. *Population Studies* 58(1):37-53.
- Ziman, J.
1984 *An Introduction to Science Studies*. Cambridge, England: Cambridge University Press.
- Zimmer, Z., A.I. Hermalin, and H. Lin
2002 Whose education counts? The added impact of adult-child relationships on physical functioning of older Taiwanese. *Journals of Gerontology: Social Sciences* 57B: S23-S32.

Zucker, L., and M. Darby

- 2005 *Socio-Economic Impact of Nanoscale Science: Initial Results and NanoBank*. (NBER Working Paper No. 11181). Available: <http://www.nber.org/papers/W11181> [accessed Nov. 2006].

APPENDIX

Biographical Sketches of Committee Members and Staff

Irwin Feller (*Chair*) is senior visiting scientist at the American Association for the Advancement of Science. He is also emeritus professor of economics at the Pennsylvania State University, where he served on the faculty for more than three decades, including 24 years as director of the Institute for Policy Research and Evaluation. His current research interests include the economics of science, the evaluation of federal and state technology programs, and the university's role in technology-based economic development. He has been a consultant to the President's Office of Science and Technology Policy, as well as numerous other government and private-sector organizations. He has served on several committees for the National Research Council, including the committee on international benchmarking of U.S. science and manufacturing modernization. He has a B.A. in economics from the City University of New York and a Ph.D. in economics from the University of Minnesota.

Wendy Baldwin is executive vice president for research at the University of Kentucky. Previously she was director of the Poverty, Youth and Gender program at the Population Council in New York. From 1993 to 2002, as deputy director for extramural research at the National Institutes of Health, Baldwin advised the director on extramural policy issues and was responsible for developing and overseeing policies and procedures for extramural research and training programs. Before that, Baldwin had a 20-year period of service to the National Institute of Child Health and Human Development. She has received many awards and distinctions, including a National Public Service Award. She has served on the American Association for the

Advancement of Science's Committee on Nominations. At the National Research Council, she served on the Panel on Adolescent Pregnancy and Childbearing. She has a bachelor's degree in sociology (*magna cum laude*) from Stetson University and master's and doctorate degrees from the University of Kentucky.

Paul B. Baltes is director of the Center of Lifespan Psychology at the Max Planck Institute for Human Development, Berlin, and professor of psychology at the Free University of Berlin. He is known for his contributions to (a) creating the field of life-span psychology, (b) the psychological study of wisdom, (c) research on cognitive aging and the plasticity of the aging mind, (d) social scenarios concerning the future of old age and an aging society, and (e) the articulation and testing of models of successful development and aging. For his work, Baltes has been honored with numerous awards, including honorary doctorates (Jyväskylä, Stockholm, Geneva) and election as a foreign member to the American Academy of Arts and Sciences and the Royal Swedish Academy of Sciences. In 2000, Baltes was elected to the Order Pour le Mérite of the Sciences and the Arts. He has a doctorate from the University of Saarland (Saarbrücken, Germany, 1967).

Richard De Veaux is professor of mathematics and statistics at Williams College. An applied statistician, he has been studying competing methods that can be used on problems in science that include artificial neural networks, as well as such techniques as trees, MARS, and MART and the application of these tools to large data sets, often called data mining. He has been a consultant for over 15 years for such companies as Hewlett-Packard, Alcoa, and other private-sector organizations. He has taught at the Wharton School and the Princeton University School of Engineering. His numerous teaching awards include a Lifetime Award for Dedication and Excellence in Teaching from the Engineering Council at Princeton. In 1998 he was elected a fellow of the American Statistical Association. He has a B.S.E. in civil engineering and an A.B. in mathematics from Princeton University and an M.A. in dance education and a Ph.D. in statistics from Stanford University.

James S. Jackson is the Daniel Katz distinguished university professor of psychology and professor of health behavior and health education in the School of Public Health at the University of Michigan, as well as director and research professor of the Institute for Social Research. He has conducted research and international, comparative studies on immigration, race and ethnic relations, physical and mental health, adult development and aging, attitudes and attitude change, and African American politics. He has served as national president of the Black Students Psychological Association and the Association of Black Psychologists. He is a member of the Institute of

Medicine. He received the 2007 James McKeen Cattell fellow award from the Association for Psychological Science. He is currently directing the most extensive social, political behavior, and health surveys on the American and Caribbean populations ever conducted. He has a Ph.D. from Wayne State University.

Janice Kiecolt-Glaser is professor of psychiatry and psychology in the Ohio State University College of Medicine, director of the Division of Health Psychology in the Department of Psychiatry, and a member of the Behavioral Medicine Research Institute. Working in the area of psychoneuroimmunology, her research, mostly in collaboration with Ronald Glaser, has demonstrated important health consequences of stress, including slower wound healing and impaired vaccine responses in older adults; in addition, their programmatic work has focused on the ways in which personal relationships influence immune and endocrine function and health. She is a member of the Institute of Medicine and a fellow of the American Psychological Association, and she received an Award for Outstanding Contributions to Health Psychology from the American Psychological Association, among other honors. She has served on the National Institute of Mental Health's Mental Health and AIDS study section. She has a bachelor's degree from the University of Oklahoma (1972) and a Ph.D. in clinical psychology from the University of Miami (1976).

Robert E. Kohler is professor of the history and sociology of science at the University of Pennsylvania. His interests include the history of science, medicine, and technology in 19th- and 20th-century Europe and America, and more recently the history and sociology of scientific practice, environmental history, environmental science, and the cultural history of science in Europe and the United States. He has been at the University of Pennsylvania since 1973. In addition to work on the culture and practices of field biology, he has also written on the history of foundations' patronage of science. He has a Ph.D. in organic chemistry from Harvard University.

Michèle Lamont is Robert I. Goldman professor of European studies and professor of sociology and African and African-American studies at Harvard University. She taught at Princeton University for 15 years before joining the Harvard faculty in 2003. She has written on the role of culture in generating social inequality; the cultural strategies of stigmatized groups for coping with racism; culture and poverty; how culture mediates the impact of discrimination on health; and many other topics. She received the C. Wright Mills Award from the Society for the Study of Social Problems. She is the current chair of the Council for European Studies, co-director of the Successful Societies Program of the Canadian Institute for Advanced Studies, and

director of the European Network on Inequality of the Multidisciplinary Program on Inequality and Social Policy at the John F. Kennedy School of Government at Harvard University. A Canadian, she has a B.A. and an M.A. in political science from Ottawa University and a Ph.D. from the Université de Paris (1983).

Leah L. Light is professor of psychology at Pitzer College. Her research interests lie in memory and aging, with a particular focus on differentiating aspects of memory that are relatively preserved in old age from those that are more affected. She has been a fellow at the Andrew Norman Institute of the Andrus Gerontology Center at the University of Southern California. Her research has been supported by grants from the National Institute on Aging, including a MERIT award, since 1981. She is a fellow of the American Psychological Association, the American Psychological Society, and the Gerontological Society of America, as well as a member of the Psychonomic Society and the Memory Disorders Research Society. She served a six-year term as editor of *Psychology and Aging* and is a past president of the adulthood and aging division of the American Psychological Association. She has a B.A. in psychology from Wellesley College and a Ph.D. from Stanford University.

Daniel McFadden is the E. Morris Cox professor of economics and director of the Econometrics Laboratory at the University of California, Berkeley. His research focuses on a variety of topics in economics and choice theory, almost all having their origins in applied problems. In recent years, his research has concentrated on deviations from the economic theory of choice, found particularly in experiments in cognitive psychology and their implications for economic analysis and the interpretation of economic data. With support from the National Institute on Aging, he has been working on the economic status of elderly Americans, looking at such questions as the adequacy of housing arrangements, financial planning, and the delivery and cost of health services. In 1981 he was elected to the National Academy of Sciences, and in 2000 he received the Nobel Prize in Economics. At the National Research Council, he has served on numerous committees, including the Committee on Applied and Theoretical Statistics and the Committee on Methods of Forecasting Demand and Supply of Doctoral Scientists and Engineers. He has a B.S. in physics (with high distinction, 1957) and a Ph.D. in behavioral science (economics, 1962) from the University of Minnesota.

Gary Sandefur is dean of the College of Letters and Science and a professor in the Department of Sociology at the University of Wisconsin–Madison. He is a social demographer known for his work on American Indians, race and ethnicity, families, poverty and inequality, and public policy. He has received

federal and state grants for research in a number of areas, including training programs for American Indians, poverty and social policy, migration, and family disruption, and was the principal investigator on a grant from the National Institute of Child Health and Human Development, Social Capital and Adolescent Well-Being, part of the NICHD Family and Child Well-Being Research Network. He served as a member of the National Research Council–Institute of Medicine Board of Children, Youth, and Families. He has a B.S. in sociology from the University of Oklahoma (1974) and a Ph.D. in sociology from Stanford University (1978).

Paul C. Stern (*Study Director*) is a senior staff officer at the National Research Council and study director of the Committee on the Human Dimensions of Global Change. His research interests include the determinants of environmentally significant behavior, particularly at the individual level, participatory processes for informing environmental decision making, and the governance of environmental resources and risks. He is the coauthor or coeditor of *Decision Making for the Environment: Social and Behavioral Science Research Priorities* (2005), *Environmental Problems and Human Behavior* (2002), *The Drama of the Commons* (2002), and *The Aging Mind: Opportunities in Cognitive Research* (2000). Stern is a fellow of the American Association for the Advancement of Science and the American Psychological Association. He has a B.A. from Amherst College and M.A. and Ph.D. degrees from Clark University, all in psychology..

Shripad Tuljapurkar is a professor of biological science and Dean and Virginia Morrison professor of population studies at Stanford University. He is also president of Mountain View Research. His research includes the dynamics and evolution of human and nonhuman populations, the sensitivity and extinction dynamics in the presence of disturbance, population aging and age structural transitions, and the evolution of senescence. He is a member of the Center for the Economics and Demography of Aging at the University of Southern California, Berkeley. He has an M.Sc. in physics from the Indian Institute of Technology, Bombay, and a Ph.D. in environmental science from Portland State University.

George E. Walker is a senior scholar at the Carnegie Foundation, where he directs the Carnegie Initiative on the Doctorate. Walker joined the foundation in January 2001; he also retains his duties at Indiana University as vice president for research and dean of the University Graduate School. Walker is active in many of the national organizations related to graduate education and research administration. Recent positions include chair of the Council of Graduate Schools (1995), chair of the Midwest Association of Graduate Schools (1996), president of the Association of American Universities As-

sociation of Graduate Schools (1997), and chair of National Association of State Universities and Land-Grant Colleges Council on Research Policy and Graduate Education (1997-1998). A theoretical physicist, he has a bachelor's degree from Wesleyan University and M.S. and Ph.D. degrees from Case Institute of Technology.

Carol Weiss is the Beatrice B. Whiting professor of education policy at the Harvard Graduate School of Education. Her research deals with evaluation, the uses of research in policy making, and the influence of ideology, interests, information, and institutional rules and structures on policy. Her two newest research studies deal with decision making by the Massachusetts State Board of Education on high-stakes testing and with federal, state, and school district decision making on drug abuse prevention education. She has served on nine panels of the National Research Council, as well as on advisory committees to the Directorate for Education and Human Resources of the National Science Foundation, the Annenberg Challenge, the World Bank, the Environmental Protection Agency, the International Development Research Center (Canada), UNESCO, the National Center for Education Statistics, the Rand Corporation, the Government Accountability Office's Program Evaluation and Methodology Division, and others. A sociologist, she has a Ph.D. from Columbia University.

David A. Wise is the Stambaugh professor of political economy at the John F. Kennedy School of Government at Harvard University. He is also the area director of Health and Retirement Programs and director of the Program on the Economics of Aging at the National Bureau of Economic Research, as well as a senior fellow at the Hoover Institution at Stanford University. He has written extensively about the saving effects of personal retirement programs—such as individual retirement accounts and 401(k) plans in the United States—and, more recently, has been evaluating the implications of the rapid spread of these programs, which now are the dominant form of saving for retirement in the United States. He is currently engaged in analysis of the retirement incentives in public social security programs around the world. He has a B.A. from the University of Washington and an M.A. in statistics and a Ph.D. in economics from the University of California, Berkeley.