Research Funding as an Investment: Can We Measure the Returns?

April 1986

NTIS order #PB86-218278

RESEARCH FUNDING AS AN INVESTMENT: CAN WE MEASURE THE RETURNS?

A TECHNICAL MEMORANDUM

Recommended Citation:

Research Funding as an Investment: Can We Measure the Returns?-A Technical Memorandum (Washington, DC: U.S. Congress, Office of Technology Assessment, OTA-TM-SET-36, April 1986).

Library of Congress Catalog Card Number 86-600525

For sale by the Superintendent of Documents
U.S. Government Printing Office, Washington, DC *20402*

Foreword

The mounting intensity of global competition in the 1980s underscores the critical role played by the basic research enterprise of the United States. Basic research is the backbone of much of the technological development that has provided not only our economic prosperity, but improvements in health, a strong national defense, and exciting and fundamental advances in knowledge.

Congress is faced with difficult decisions regarding funding for research. The House Science Policy Task Force asked OTA to provide information on the extent to which decisionmaking would be improved through the use of quantitative mechanisms associated with the concept of investment. If investing in science is similar to investment in the financial sense, can the returns be meaningfully predicted and measured? Can reasonable investment criteria be devised? OTA concluded that while there are some quantitative techniques that may be of use to Congress in evaluating specific areas of research, basic science is not amenable to the type of economic analysis that might be used for applied research or product development. OTA also concluded that even in the business community, decisions about research are much more the result of open communication followed by judgment than the result of quantification.

Much of the vitality of the American research system lies in its complex and pluralistic nature. Scientists, citizens, administrators, and Members of Congress all play various roles leading to final decisions on funding. While there may be ways to improve the overall process, reliance on economic quantitative methods is not promising. Expert analysis, openness, experience, and considered judgment are better tools.

John H. Gibbons

JOHN H. GIBBONS
Director

Workshop Participants

Harvey Averch
National Science Foundation

Radford Byerly, Jr.
U.S. Congress
House Committee on Science and Technology

Daryl Chubin
Georgia Institute of Technology

Henry Hertzfeld
Consultant

John Holmfeld
U.S. Congress
House Committee on Science and Technology

David Mowery
Carnegie-Mellon Institute

J. David Roessner
Georgia Institute of Technology

Nestor Terleckyj
Center for Socio-Economic Analyses
National Planning Association

Reviewers and Other Contributors

William Alumkal
E.I, du Pont de Nemours & Co.

Nicholas Ashford
Massachusetts Institute of Technology

Alden Bean
Lehigh University

Edwin L. Behrens
Proctor & Gamble

Harvey Brooks
Harvard University

Charles Buffalano
Defense Advanced Research Projects Agency

H. Roberts Coward
Center for Research Planning

Susan Cozzens
National Science Foundation

Gordon Enk
International Paper Co.

Daniel Fink
Consultant

James Fletcher
Consultant

J. Jeffrey Franklin
Center for Research Planning

Howard Gobstein
General Accounting Office

Zvi Griliches
Harvard University

Walter Hahn
George Washington University

Christopher Hill
Congressional Research Service

Noel Hinners
National Aeronautics and Space Administration

Allan R. Hoffman
National Academy of Sciences

Benjamin Huberman
Consultants International

Bill Kirsch
University City Science Center

Genevieve Knezo
Congressional Research Service

Charles F. Larson
Industrial Research Institute

Fred Leavitt
Dow Chemical, USA

John M. Logsdon
George Washington University

Alexander MacLachlan
E.I. du Pont de Nemours & Co,

L.M. Magner
E.I. du Pont de Nemours & Co.

Edwin Mansfield
University of Pennsylvania

Bruce Mazlish
Massachusetts Institute of Technology

Norman Metzger
National Academy of Sciences

Robert P. Morgan
Washington University

Nathan Rosenberg
Stanford University

Mark Ross
University of Michigan

Charles F. Sabel
Massachusetts Institute of Technology

D.M. Von Schriltz
E.I. du Pont de Nemours & Co.

Ora Smith
Office of Science and Technology Policy

Elly Thomas
National Science Foundation

John N. Warfield
George Mason University

Robert Weaver
Pennsylvania State University

Philip Webre
Congressional Budget Office

Research Funding as an Investment: Can We Measure the Returns?
OTA Project Staff

John Andelin, *Assistant Director, OTA*
Science, *Information and Natural Resources Division*

Nancy Carson Naismith
Science, Education, and Transportation Program Manager

Project Staff

Eugene Frankel, *Project Director (After* September 1985)

Barry J. Holt, *Project Director* (Through September 1985)

Kathi Hanna Mesirow, *Analyst*

Lisa Heinz, *Analyst*

Mia Zuckerkandel, *Research Assistant*

Marsha Fenn, *Administrative Assistant*

Betty Jo Tatum, *Secretary*

Chris Clary, *Secretary*

Gala Adams, *Clerical Assistant*

Contractors

Evan Berman Daryl Chubin Kevin Finneran
Henry Hertzfeld David Mowery J. David Roessner

Other OTA Contributors

John Alic Philip P. Chandler II
Marcel C. LaFollette Martha Harris

Contents

List of Tables

List of Figures

Chapter 1

Executive Summary: Overview and Findings

Chapter 1

Executive Summary: Overview and Findings

The assumption that federally funded scientific research leads to economic benefits for the country has been fundamental to government science policy since the end of World War II. Analysts have abandoned the linear model that sees a simple progression from basic research to applied research to product development, but they still believe that scientific research plays a vital role in technological progress and consequently in economic growth. Economic returns, however, are neither the sole nor the primary purpose for Federal research spending. The advancement of knowledge and specific mission agency goals such as national security, public health, and the exploration of space are all essential parts of the rationale for Federal research spending.

Several trends have combined in recent years to make some policymakers more interested in economic and other quantifiable measures of research success and benefits. Technology is becoming an essential component of economic competitiveness; Federal budget constraints are forcing lawmakers to reevaluate spending and to look for ways to compare the value of widely divergent government programs; quantification of program success offers the hope for an objective measure that could simplify politically contentious decisions about increasingly esoteric and complex scientific research. One approach to simplifying research evaluation is to view Federal research spending as an investment that should produce a measurable economic return.

The Task Force on Science Policy of the House Committee on Science and Technology has raised the issue of whether the metaphor of research funding as an investment can be used as a practical aid to Federal research decisionmaking. "Can Federal funding for science be viewed as an investment and be measured in a way comparable to other forms of economic investment?" the Committee asked in its Report *on a Study of Science Policy*. Specifically, the Committee asked O-J-A to study "the models and other analytical tools developed by economists to judge capital investments, and the applicability and use of these models and tools to government funding of scientific research. "

To carry out this study, OTA conducted a comprehensive search of the literature on the economic returns to investment in scientific research, met with numerous economists and public policy analysts who have studied this issue, conducted interviews with research decisionmakers in industry and in government, and carried out in-depth studies of the quantitative methods available to evaluate the progress of scientific research. This technical memorandum presents the findings of that investigation.

ECONOMIC RETURNS

Economists have shown a strong positive correlation between research and development (R&D) spending and economic growth. They have estimated private returns in excess of 20 percent per year and social returns in excess of 40 percent on private sector R&D expenditures. They have *not* been able to show comparable returns, and at times been unable to show *any* returns, on Federal R&D expenditures, except for some applied research programs in agriculture, aeronautics, and energy designed to improve industrial productivity. These findings are discussed at length in chapter 2.

The economists who have carried out these studies point out a number of reasons why economic return on investment calculations may be inappropriate for evaluating government R&D expenditures. First, the return-on-R&D-investment studies carried out to date measure an average return on a total previous investment. They give little guidance as to the marginal return that can be expected from the next incremental investment in R&D, which is the decision that policymakers must make. Second, most government expenditures, including R&D expenditures, are for so-called "public goods" whose market value is, by definition, extremely difficult to measure in economic terms. Third, despite the success of retrospective studies, there are no reliable formulae to relate future R&D expenditures to productivity improvements or other economic benefits. Predictions of future returns on investment cannot be made without such relationships.

Financial counselors and economists have developed techniques for selecting investments in situations involving risk and uncertainty. These techniques include capital investment methodologies, portfolio analysis, and financial investment' models. All of these techniques have heuristic properties that could guide investment in research and development. For example, spreading the risk among a number of projects is one response economists sometimes recommend in cases involving great uncertainty. However, the formal models themselves are not especially helpful to research decisionmaking. They assume that the decisionmaker can estimate in dollar values the benefits from potential investments and know or estimate the probability of achieving those benefits. Neither of these assumptions is applicable to government-funded research, except in special cases. The principal benefit of research, especially basic research, is new and often unexpected knowledge, which cannot be assigned a direct economic value.

Investment models also assume that the benefits of the investment return to, or are appropriable by, the investor. New knowledge-by contrast—is available to anyone to use. This is one of the reasons basic research is considered a public good, requiring government support.

Research leads to productivity improvements and economic growth primarily through technological innovation. However, the relationship between research and innovation can be long-term, indirect, and unpredictable. Studies of technological innovations have shown them to depend on research results that are decades old and often in seemingly unrelated fields. Moreover, the transformation of research into economically success" ful innovation depends on factors in the economy that are completely outside the research process. These factors include the climate for investment; government tax, regulatory, and patent policy; the degree of competitiveness and entrepreneurialism in industry; the state of the capital markets; foreign competition; and wages, unionization, and other characteristics of the work force. A highly successful basic research effort may never generate technological innovation or economic payoff if other factors in the economy are not conducive to technological change.

Some observer: argue that if economic returns are to be the primary measure of our research effort, we should focus our attention, as a Nation, on the factors that link science to technology and innovation. The United States spends less of its R&D budget than West Germany, France, England, or Japan on research related directly to industrial productivity. Efforts to improve that situation could include an increased emphasis on technology transfer, increased support for generic research related to industrial needs, and adaptation of more focused forecasting and planning for industry-related R&D. (See ch. 5.)

Applied research, whose goal is the solution of practical problems, can be more closely associated with economic activity. However, most of the applied research in the Federal Government is carried out by agencies whose mission objectives—defense, health, space—are not readily quantifiable in dollar terms. Table 1 shows the estimated 1985 Federal basic and applied research budgets by agency. As can be seen, in applied research the Departments of Defense (DOD) and Health and Human Services (DHHS) were the two largest contributors, with the Department of Energy (DOE) and the National Aeronautics and Space Administration (NASA) third and fourth. All of the DOD and DHHS applied research re-

Table l.—Federal Obligations for Research and Development by Character of Work and R&D Plant: Fiscal Years 1984-85 (thousands of dollars)

			Research			
Fiscal year and agency	Total R&D and R&D plant	Total R&D	Basic research	Applied research	Development	R&D plant
Fiscal *year* 1984 *(estimated):*						
Total, all agencies	46,554,924	44,835,777	6,981,031	8,127,270	29,727,478	1,719,145
Department of Agriculture	925,364	871,942	386,442	455,594	29,906	53,422
Department of Commerce	367.252	360.021	20.522	272,644	66.855	7,231
Department of Defense	27,987,145	27,540,045	816,590	2,168,184	24,555,271	447,100
Department of Energy[a]	5,770,604	4,825,576	841,671	1,231,733	2,752,172	945,028
Department of Health and Human Services[b]	4,921,924	4,864,292	2,793,052	1,705,911	365,329	57,632
Department of the Interior	427,558	421,825	124,667	276,330	20,828	5,731
Department of Transportation	538,429	515,929	600	81,990	433,339	22,500
National Aeronautics and Space Administration .,	3,044,400	2,888,900	689,133	1,012,031	1,187,738	155,500
National Science Foundation	1,247,580	1,238,480	1,172,466	66,014	—	9,100
Veterans Administration	228,100	220,900	15,200	189,700	16,000	7,200
Other agencies	1,096,568	1,087,867	120,688	667,139	300,040	8,701
Fiscal *year* 1985 (estimated):						
Total, all agencies	54,072,393	52,253,607	7,637,587	8,396,633	36,219,387	1,818, 786
Department of Agriculture	926,711	898,941	419,727	449,981	29,233	27,770
Department of Commerce	282,357	270,559	18,416	201,187	50,956	11,798
Department of Defense , .	34,510,984	34,142,084	913,195	2,408,204	30,822.685	368,900
Department of Energy[a]	6,146,700	4,962,272	944517	1,268,964	2,748,791	1 184,428
Department of Health and Human Services[b]	4,967,872	4,953,972	2,925,916	1,679,147	348,909	13,900
Department of the Interior	369,209	368,989	102,762	248,556	17,671	220
Department of Transportation	505,704	495,204	400	79,630	415,174	10,500
National Aeronautics and Space Administration	3,499,400	3,339,400	826,721	1.088,063	1 424,616	160,000
National Science Foundation	1,426,567	1,414,017	1,335,809	78.208	—	12550
Veterans Administration	207,600	194,500	15,000	160,000	19,500	13,100
Other aqencies	1,229,289	1,213,669	135,124	736.693	341,852	15,620

[a]Data shown for fiscal years 1956-73 and fiscal years 1974-76 represent obligations of the Atomic Energy Commission (A EC) and the Energy Research and Development Administration. respectively
[b]Data shown for fiscal years 1955-713 represent obligations of the Department of Health, Education, and Welfare

SOURCE: National Science Foundation

lates to national defense and health, two public goods that are not readily measured in economic terms. With the exception of approximately $200 million in aeronautics research, all of NASA's applied research relates to its space activities, which are not primarily designed to produce economic payoffs. Only the DOE, Department of "Agriculture, Department of the Interior, NASA Aeronautics, and Department of Commerce applied research programs have primary objectives related to improving the economic performance of an industry or the economy as a whole. In sum, nearly two-thirds of the Federal applied research budget is related to the production of public goods, whose primary value is not measured in economic terms.

Table 1 reveals another barrier to the use of economic models for research investment in the Federal Government-decentralized planning and decisionmaking. Six Federal agencies have a share of the total Federal research budget in excess of 5 percent: DHHS, DOD, DOE, NASA, the National Science Foundation (NSF), and the Department of Agriculture in descending order. The Office of Management and Budget, which could develop an overall national research budget, is divided into functional directorates that each examine only part of that budget. The Office of Science and Technology Policy does consider the Federal research budget as a whole, but it has no decisionmaking authority over Administration budget requests.

In Congress, responsibilities are equally dispersed. Three different authorizing committees and six largely independent appropriations subcommittees scrutinize the Federal R&D budget in each House. Thus even if some economic or financial model could be devised to determine the return on the Federal research "investment" and serve as a guide to allocating scarce resources, there is no single decisionmaker in the Federal Government who could ensure its uniform application across all research fields and budgets. Without such a decisionmaker, such a model would have little operational power or efficacy.

BIBLIOMETRICS AND SCIENCE INDICATORS

A major problem with the use of economic models in research decisionmaking is that they deal with economic "indicators" that are at best indirectly influenced by research. To measure research output more directly, alternative "indicators" have been extensively developed by students of science policy over the past two decades. However, these have only recently begun to be considered seriously by research policymakers as possible aids to their decision processes. The two main approaches are bibliometrics, which evaluates research output via scientific publications; and science indicators, which measure the vitality of the research enterprise in terms of degrees, personnel, awards, and education. Although these methods appear to be more appropriate measures of scientific quality and productivity, they do not offer the decisionmaker the simple, quantitative economic "bottom line" that economic models provide. The "indicators" can only supplement, and not replace, informed peer judgment of the scientific process. But they can help complete the anecdotal, fragmentary, and, necessarily, somewhat self-interested picture of the state of science presented by the researchers themselves. Science indicators, and especially bibliometric measures, are reviewed in chapter 3 of this technical memorandum.

Bibliometrics is based on the assumption that progress in science comes from the exchange of research findings, and that the published scientific literature is a good indicator of a scientist's knowledge output. Publications are the medium of formal information exchange in science and the means by which scientists stake their claims to intellectual "property. " Therefore, the more publications a scientist has, the greater is his or her presumed contribution to knowledge.

Simple publication counts have a number of obvious flaws: quantity of publications does not measure the quality of the knowledge contained therein; publications vary greatly in creativity and impact. Simple counts also cannot be used for cross-disciplinary analysis because of differences in publication rates by research field, type of research, research institutions, and a number of other external factors.

Citation analysis addresses the problem of measuring the quality of research output. It assumes that the greater the quality, influence, or importance of a particular publication, the more frequently it will be cited in the scientific literature. Citation counts based on comprehensive databases are being used on a limited basis to monitor the performance of research programs, facilities and faculties in Europe, and at the National Institutes of Health (NIH) and NSF.

The problems of citation analysis include: technical problems with the database, variations in the citation rate over the "life" of a paper, the treatment of critical or even refutational citations, variations in the citation rate with the type of paper, and biases introduced by "self-citation" and "in-house" citations. Developers of the citations database are working to minimize these problems, but some are inherent. Sophisticated variations on the approach include co-citation and co-word analysis, which are described in chapter 3,

Combinations of several research productivity indicators (publications, citation counts, and peer evaluation) have been used in the hope of overcoming problems associated with each method en its own. To the extent that the "partial indicators" converge, proponents argue, the evaluation may be more meaningful than if only one indicator were used. Significant degrees of convergence have been found by using this methodology to evaluate large physics and astronomy facilities. Since partial indicators depend in large part on a peer review system, they can be used as an independent check on scientists' peer assessments of research activities.

Despite the limitations of bibliometrics, NSF and NIH have undertaken extensive studies to refine the techniques and explore their applicability to research program evaluation. In addition, agencies of the French, Dutch, and British Governments, and the European Organization for Economic Cooperation and Development have applied some of these indicators to research programs in their countries. The results of these studies, and the limitations of this methodology, are discussed in chapter 3.

Science "indicators" assess the ongoing vitality of the research enterprise, complementing the "output" measures of bibliometric analyses. These indicators include statistics on scientific and engineering personnel; graduate students and degree recipients by field, sector, and institution; and the support for graduate education and training. NSF, NIH, and the National Research Council. (NRC) publish detailed indicators on a regular basis. However, the science policy community lacks consensus on which indicators are most useful or reliable. A report or workshop on the use of science indicators to measure the health of the research effort in the United States would be a useful first step in that direction.

It is important to remember that all measures or "indicators" of research inevitably are flawed. Any number describing research is an abstract symbol that depicts, imperfectly, only one aspect of it. Choosing one measure over another implies that the measurement user has made some assumption about what is important. The chosen measure has meaning only through interpretation.

These points underscore the subjective nature of quantitative measures of research—''objectivity" is only apparent. Attaching numbers to some phenomena allows the expression of certain features in symbols that can be manipulated and configured for analysis. This ability is invaluable for analytical comparisons and the descriptions of trends. Nevertheless, a number remains no more than an abstract symbol that someone decided best captures a particular aspect of some real-world phenomena.

RESEARCH DECISIONMAKING IN INDUSTRY AND GOVERNMENT

To determine the degree to which economic and noneconomic techniques are used by research managers today, OTA reviewed the literature on the use of these techniques in industry and government and interviewed experienced officials in both sectors. A list of those techniques is provided in table 2. The findings were quite surprising.

In industry, where one might expect quantitative techniques to prevail due to the existence of a well-defined economic objective for the individual firm or business, OTA found great skepticism among research managers about the utility of such techniques. Managers found them to be overly simplistic, inaccurate, misleading, and subject to serious misinterpretation. At the project selection and program evaluation levels, there is little systematic data about the use of quantitative techniques. Most articles describe a process adopted by one firm or another without any indication as to how widespread the practice is in industry as a whole. This literature is reviewed in chapter 4.

Peer review dominates program evaluation in industry, with an occasional firm attempting bibliometric analyses. For project selection, firms use standard economic return on investment techniques for projects at the development end of the cycle, where costs and benefits are generally well known and the risk can be quantified using past experience. At the basic research end of the spectrum, industry's project selection techniques tend to be quite subjective and informal, supplemented occasionally by scoring models. (See ch. 4 for definitions of the different techniques used in project selection.) At the applied research or exploratory development stage, simple, unsophisticated selection procedures, based on a page or two of qualitative information or a simple rating scheme, dominate.

Table 2.—Quantitative Methods Used To Evaluate R&D Funding

Economlc (measure output in terms of dollars or productivity)
. Macroeconomic production function (macroeconomic)
• Investment analysis
–Return on investment (ROI)
–Cost/benefit **analysis (CBA)**
—Rate of return
—Business opportunity
• Consumer and producer surplus
Output (measure output in terms of published information)
• Bibliometric (publication count, citation, and co-citation analysis)
• Patent count and analysis
• Converging partial indicators
• Science indicators
Project selection models
. Scoring models
• Economic models
• Portfolio analysis (constrained optimization)
. Risk analysis and decision analysis

SOURCE: Office of Technology Assessment.

At the level of strategic planning and resource allocation for R&D in industry, some interesting patterns have emerged. Industry tended to fund research somewhat unquestioningly in the 1950s and 1960s, only to become skeptical of a lack of demonstrable return on the investment in the 1970s. Each industry tended to have a rule of thumb; R&D should bes percent of sales, or perhaps 10 percent in an R&D-intensive industry. In the 1970s, corporate strategic planning came into vogue, and technological change came to be recognized as an integral part of corporate planning. R&D planning and budgeting was integrated into the overall corporate strategic effort. Many firms set up committees and other formal mechanisms to assess long-term technical opportunities, establish broad goals for the commitment of resources, ensure that resources are properly allocated to develop the technology necessary to support those goals, approve major new product programs, and monitor progress.

The primary goal of such committees appears to have been to ensure that R&D managers communicate regularly and formally with planning, financing, marketing, manufacturing, and other concerned parts of the corporation in setting and achieving technological goals. Corporate managers have learned that R&D planning and budgeting is primarily an information and communication process, involving many persons and many levels of the corporate hierarchy, using many criteria and several iterations. The goal of corporate managers has been to improve communication by involving all affected parties. Economic and financial modeling appear to play a secondary role in this process, seining primarily as inputs to overall corporate strategic planning.

Government R&D managers also avoid quantitative techniques for project selection and program evaluation. Surveys of government research managers reveal little use of quantitative methods for choosing projects or evaluating program: (some notable exceptions are discussed inch. s) Peer review tends to be the preferred method of project selection, with the term "peer" often broadened to include agency technical staff. Bibliometric techniques have been used extensive> by NIH and on a limited basis by NSF in program evaluation. NASA and the National Bureau of Standards have carried out economic return-on investment analyses, with limited success.

Budgeting for research and development share many of the characteristics of traditional Federa budgeting. It is incremental, fragmented, special ized, repetitive, and based, to a large degree, or recent history and experience. Much of the bud geting is carried out by experts in narrow special ties, who focus their attention on increments to existing base programs. Attempts to "rationalize the system by introducing techniques such a "program planning and budgeting (PPB)" and "zero based budgeting (ZBB)" have largely bee: abandoned as unworkable and inappropriate given the political nature of the Federal budge process.

Some R&D forecasting and strategic planning is carried out by agency advisory committees such as NSF's National Science Board, DOE Energy Research Advisory Board, and NHI's advisory councils. NRC and its constituent bodies forreally review Federal research programs and produce *Research Briefings* and *Five-Year Outlook* that identify promising new avenues of research None of those efforts constitutes true strategic planning or forecasting. The Japanese, however provide an interesting model of systematic forecasting and planning for R&D.

Chapter 5 also describes the R&D forecasting carried out by Japan's Science and Technology Agency and Ministry of International Trade and Industry. These forecasts identify research areas of long-term strategic importance using background information on research trends gleaned from industry, government, and academic reports from around the world. They incorporate "technology-push" and "market-pull" perspectives by involving both laboratory researchers and industrial users, and utilize a bottom-up rather than a topdown approach, drawing heavily on recommendations from the affected communities. The process provides a forum for people from different groups and different professions to communicate about R&D priorities. This enables policymakers, professional forecasters, scientific analysts, and academic and industrial researchers to coordinate research plans and to form a consensus on priorities for future strategic research. Participants have a stake in the successful outcome and follow-through of the process, which tends to make the forecasts self-fulfilling. The emphasis on communication and involvement of all affected parties is strikingly similar to the lessons learned by U.S. corporate management with respect to R&D planning and resource allocation described above.

The review of industry and government R&D decisionmaking presented in chapters 4 and 5 leads to two conclusions. First, R&D management and resource allocation are complex decision-making processes involving trade-offs between factors that often cannot be precisely measured or quantified. Any effort to substitute formalistic quantitative models for the judgment of mature, experienced managers can reduce rather than improve the quality of R&D decisionmaking. The resistance of R&D managers to the use of quantitative decision tools is, to some degree, a rational response to the complexity and uncertainty of the process.

Second, the process of decisionmaking can often be as important as the outcome. In both the U.S. and Japanese cases, bringing together experts from a variety of fields and sectors and providing them with a vehicle to discuss R&O priorities, budgets, and plans, was critical to success. It may be that discussions of R&D resource allocations in the United States should focus less on the overall numbers and more on the process by which those numbers were generated, with special attention paid to questions of stakeholder involvement and communications.

SUMMARY

In summary, OTA finds that the metaphor of research funding as an investment, while valid conceptually, does not provide a useful practical guide to improving Federal research decisionmaking. The factors that need to be taken into account in research planning, budgeting, resource allocation, and evaluation are too complex and subjective; the payoffs too diverse and incommensurable; and the institutional barriers too formidable to allow quantitative models to take the place of mature, informed judgment. Bibliometric and other science indicators can be of some assistance, especially in research program evaluation, and should be used more widely. However, they are extremely limited in their applicability to inter-field comparisons and future planning. The research planning and budgeting experience in some U.S. corporations and the R&D forecasting efforts in Japan suggest a need to improve communication between the parties that carry out and utilize research, and to assure that a wide range of stakeholders, points of view, and sources of information are taken into account in formulating R&D plans and budgets.

Chapter 2

Measuring the Economic Return: Progress and Problems

Chapter 2

Measuring the Economic Returns: Progress and Problems

INTRODUCTION

This chapter discusses past and current use of economic measures of the return on research and development (R&D) as an investment in individual industries and at the national level. Different approaches distinguish between direct and indirect returns, private and Federal spending, basic and applied R&D. This chapter examines the difficulties and ambiguities encountered when trying to extend the analysis of private sector R&D spending to Federal R&D investments. It reviews attempts— of mixed success-to use econometric methods to measure the returns on Federal R&D dollars in three industries that have been well-studied: agriculture, aviation, and health.

Federal R&D dollars may also have indirect effects on productivity by triggering spinoffs or spillovers; the chapter looks at the National Aeronautics and Space Administration (NASA) as an example of the use of econometric methods to measure such indirect effects. The chapter concludes that while econometric methods have been useful to track private R&D investment within industries, the methods fail to produce consistent and useful results when applied to Federal R&D support. From these findings, OTA concludes that economic investment models are not likely to be of great utility in helping to guide Federal research decisionmaking .

ECONOMETRIC STUDIES AT THE NATIONAL LEVEL

Over the past three decades economists have attempted to investigate the effect of research expenditures, technological change, and other research-related inputs to production on the growth of GNP, productivity, and employment. Their basic production function approach has been to separate the inputs to the economy into three groups: capital and labor supply—the major factors determining the productivity of a firm, industry, or national economy—and an "other factors" category, assumed to account for all changes in productivity that could not be explained by changes in labor and capital. This residuals category includes scientific knowledge, technological advance, managerial and marketing expertise, economies of scale, the health and education of the work force, and other factors that affect the efficiency of resource use.

In the 1950s, economists recognized that residual factors were a major influence in economic growth. Using the "factor productivity method, " E.F. Denisen attributed 20 percent of the growth in the Nation's real income between 1929 and 1957 to "advance of knowledge" and 11 percent to economies of scale.[1] Such studies indicated that technological change, however defined, is important to national economic performance. [2]

Private R&D Investment

Building 0,1 these studies, in the late 19 SOS economists began to include R&D expenditures (assumed to be a rough indicator of technological advance) as an input to their productivity calculations, along with capital and labor. Numerous studies found a strong correlation between R&D spending and productivity growth. Looking at R&D as an investment, economists sought

[1] E.F. Denisen, *The Sources of Economic Growth in the U.S* (New York: National Bureau of Economic Research, 1962).

[2] Zvi Griliches, "Issues m Assessing the Contribution of Research and Development to Productivity Growth *The Bell Journal of Economics*, *vol. 10*, spring 1979.

to measure its rate of return. Fellner[3] calculated a 31 to 55 percent rate of return for the entire economy. Terleckyj[4] estimated a 29 percent return to firm-financed R&D. Mansfield[5] estimated a 40 to 60 percent return in the chemical industry and Link* estimated 21 percent in the petroleum industry. More recent in-depth studies confirm the correlation between private R&D spending and productivity increases (see box A).

These studies are representative of the strong and consistently positive correlation found between privately financed R&D and productivity growth in the manufacturing industries. They suggest that econometric analysis of private R&D spending produces estimates useful in evaluation and planning. However, the wide range of calculated rates of return to R&D spending and the inability to assign causality to the correlations reflect the tentative and hypothetical nature of the methodologies. Each study works with different assumptions and definitions. Results are most definitive and consistent for private spending within one firm or industry, where it is easiest to define and measure inputs and outputs.

Social return often exceeds the private rate of return, as a company doing the R&D cannot reap all the benefits from its work. One industry's R&D can spin off substantial benefits to other industries and other sectors of society, a difficult output to quantify. In studies by Mansfield and others, the social rate of return was two or more times the private rate of return.[7]

In examining the applications of these economic models it should be kept in mind that they are only hypothetical constructs that attempt to describe complex events. Zvi Griliches, one of the foremost users of these models, warns that the equations in the models reveal correlation, not causality.[8] Nor do the models reveal the path by which R&D investment allegedly leads to productivity improvements. Moreover, the need to treat R&D as the "residual," or "the thing that remains after everything else is accounted for, " further weakens the proof of relationship, since it is entirely possible that other components of the residual exist, but have not been included in the analysis. Finally, the production function approach of neoclassical economics is simply an hypothesis about the way the world works; it has not been proven that such production functions exist or take the form assumed by economists. For all these reasons the impressive returns on private sector R&D investment reported above should be viewed with caution.

The Returns, or Lack Thereof, to Federally Funded R&D in Specific Industries

Econometric approaches have been unsuccessful in establishing a return on federally funded R&D. Unlike the strong and consistently positive correlations found between privately financed R&D and productivity growth in the manufacturing industries, only weak and inconsistent correlations have been found for federally funded R&D. Terleckyj, in the 1975 study reported above, found that for the 20 manufacturing industries he studied, "the coefficients for government-financed R&D are not statistically significant, and the coefficient for government-financed R&D performed in industry is actually negative. "[9] A decade later Terleckyj reported subsequent studies that confirmed the weak indicators and smaller effects of government-funded R&D.[10] Even in two industrial sectors enjoying high, long-term government funding and interest—aircraft manufacturing, and communication and electronic components—Ter-

[3]W. Fellner 'Trends in the Activities Generating Technological Progress, " *American Economic Review, vol. 60,* March 1970, pp. 1-29.

[4]Nestor E. Terleckyj, *Effects of R&D on the Productivity Growth of Industries: An Exploratory Study* (Washington, DC: National Planning Association, 1974).

[5]Edwin Mansfield, "Rates of Return From Industrial Research and Development, " American *Economic Review, vol. 55,* May 196.S, pp. 310-322.

[6]A.N. Link, "Productivity Growth, Environmental Regulations and the Composition of R& D," *The Bell Journal of Economics, vol. 13,* autumn 1982, pp. *166-l@.*

[7]Edwin Mansfield, "The Economics of Innovation, " *Innovation and U.S. Research,* W. Novis Smith and Charles F. Larson (eds.), ACS Symposium Series *129,* Washington, DC, 1980, pp. *96-97.*

[8]Ibid., p. *24,* emphasis added.

[9]Nester E. Terleckyj (cd.), *State of Science and Research: Some New Indicators* (Boulder, CO: Westview Press, 1977), p. 131.

[10]Nester E. Terleckyj, "Measuring Economic Effects of Federal R&D Expenditures: Recent History With Special Emphasis on Federal R&D Performed in Industry, " paper presented to the National Academy of Sciences Workshop on "The Federal Role in Research and Development, " Nov. *21-22, 1985,* p. *5.*

Box A.—Recent Research Productivity Studies

Nestor Terleckyj, studying the productivity of entire industries in 1974, found that an industry's rate of productivity increase is directly related to the amount of its own R&D and to the amount of R&D carried out by its supplier industries.[1] In a study of the relationship between total factor productivity and R&D in 33 manufacturing and nonmanufacturing industries between 1948 and 1966, Terleckyj estimated a 28 percent productivity return on private R&D investment in the manufacturing industries. He found an even higher implicit productivity return on company-sponsored R&D by taking into account the R&D inherent in purchases from supplier industries. For the nonmanufacturing industries the correlation was much weaker, and in some cases actually negative.[2]

Zvi Griliches, in a 1975 study of 883 companies representing more than 80 percent of all the industrial R&D conducted in the United States, found a .7 percent rate of return to total R&D, private plus government funded, for the period 1957-65. There was a wide range in the rate of return by industry, with the chemical industry at the top at 93 percent; electric equipment and aircraft and missiles at the bottom at 3 to 5 percent; and metals, machinery, and motor vehicles in the middle at 23 to 25 percent. For privately financed R&D alone, Griliches found a substantially higher average return of 32 to 40 percent.[3] Terleckyj found this return to be quite comparable to his own value for the manufacturing industries of 37 percent return on private R&D when only direct R&D inputs were considered.[4]

Griliches, in a follow-up to his 1975 study, found that firms that spend a larger fraction of their R&D on basic research are more productive.[5] He found that basic research had 2.5 to 4.5 times as great an effect on productivity per dollar invested as total R&D. However, he cautioned that R&D or basic research may not drive productivity and profitability successes, but the correlation could well be that "success allows firms to indulge in these types of luxury pursuits."

Edwin Mansfield, the third major analyst in this field, refined Terleckyj's work on the 20 manufacturing industries by dividing R&D into its basic and applied components. He found a "strong relationship between the amount of basic research carried out by an industry and the industry's rate of productivity increase during 1948-1966."[6] In a further study of 37 innovations Mansfield compared the return on R&D for those innovations to the firm making the investment (the "private return") with the return to society as a whole (the "social return"). He found a median private rate of return of about 25 percent, but a median social return of close to 70 percent.[7]

[1] Edwin Mansfield, "Research and Development, Productivity, and Inflation," Science, vol. 209, Sept. 5, 1960, p. 1,091.

[2] Nestor E. Terleckyj (cd.), "Estimates of the Direct and Indirect Effects of Industrial R&D on Economic Growth," *The State Of Science and Research: Some New Indicators* (Boulder, CO: Westview Press, 1977), pp. 125-132.

[3] Ibid., pp. 133134.

[4] Ibid., p. 134.

[5] Zvi Griliches, "Productivity, R&D, and Basic Research at the Firm Level in the 1970s," NBER Working Paper No. 1S47, typescript (National Bureau of Economic Affairs, 10s0 Massachusetts Avenue, Cambridge, MA 02138), January 1985, p. 16.

[6] Edwin Mansfield, "Basic Research and Productivity Increase in Manufacturing," *American Economic Review*, vol. 60, No 5, December 1980, p. 866

[7] Edwin Mansfield, "How Economists See R& D," *Research management*, July 1982, p. 27.

leckyj "found strong positive association between private R&D and productivity," but "no effect of *government* R&D."[11]

Measuring Federal R&D spending is more complex than in the private sector. Tracing outputs through the long and nebulous path from basic research to commercial product is especially difficult. A company does research aimed at a specific product or market, controls the entire product development process, manages its marketing, and has a clear record of inputs and outputs. Federal research managers do not target R&D so sharply, have virtually no say in private sector decisions to develop a product, and have no influence and often no knowledge of what is happening in the market.

Terleckyj attributes the failure to find a return on federally sponsored industrial R&D to the fact

[11] Ibid., p. 6.

that "government funded industrial R&D is a public good and therefore is used by all users to the extent where its marginal product is zero." Therefore, according to Terleckyj, "its contribution to productivity cannot be observed statistically by traditional techniques and approaches. "[12] In addition:

> there is an inherent conceptual limitation in the national income accounting (and the GNP data) in that it attempts to measure the real cost and the real product of the public sector at the same time. While the resource cost utilized in the public sector can be identified, the real output of public goods cannot be measured because their marginal product and implicit price is always zero.[13]

The inability to find a meaningful correlation between government-funded R&D and productivity increases in the economy as a whole has led economists to examine more closely the indirect impact of Federal R&D on privately funded industrial R&D. According to Terleckyj, studies done in the past 6 years "indicate that in most cases government R&D expenditures have been positively related to private R&D expenditures. "[14] Peter Reiss, reviewing the same literature, reports a "general impression that Federal R&D is a complement to private R&D efforts, " but finds a lack of "very good conceptual models of how Federal R&D affects private R&D incentives. "[15]

Frank Lichtenberg has attempted to distinguish the direct and indirect links between Federal arid private R&D. He argues that Federal R&D expenditures "may, in principle, increase the average and marginal cost of private R&D performance by driving up the prices of R&D inputs''—or "crowding out" private R&D. Alternatively, federally sponsored R&D may leverage private R&D, reducing the costs of private research and innovation and raising the productivity of private R&D—the "spillover effect." He finds econometric evidence for the crowding out hypothesis in the short run, although less so in the long run. He finds limited evidence for cost-reducing spillovers but concludes that "it is probably the case that a small fraction of federally supported R&D generates very large spillovers (some of which may be negative.)''"

[12]Ibid., p. 7.
[13]Ibid., p. 8.
[14]Ibid., p. 9.

"Peter C. Reiss, "Economic Measures of the Returns to Federal R& D," paper presented at the National Academy of Sciences Workshop on 'The Federal Role in Research and Development, " Nov. 21-22, 1985, pp. 11-12.

'*Frank R. Lichtenberg, "Assessing the Impact of Federal Industrial R&D Expenditures on Private R&D Activity, " paper presented to the National Academy of Sciences Workshop on "The Federal Role in Research and Development, " Nov. ***21-22, 1985,*** pp. **31-32.**

INDUSTRY CASE STUDIES: AGRICULTURE, AVIATION, AND HEALTH

Despite the problems in linking government R&D expenditures to productivity improvements in the economy as a whole, studies have shown sector-specific productivity improvements from targeted government R&D programs. This section looks at econometric analyses of Federal R&D support of three industries-agriculture, aviation, and health—whose long and heavy dependence on Federal R&D financing has made it feasible for economists to estimate inputs, outputs, and rates of return. The results of, and problems with, those evaluations are presented below.

Agriculture

For nearly a century, since the passage of the Hatch Agricultural Experiment Station Act in 1887, the Federal Government has had a program to support applied research related to improved farm productivity. The program today has three main elements: 1) the Agricultural Research Service, which funds research and technology transfer projects at the USDA's own research stations and at state universities; 2) the Cooperative Research Service, which consists primarily of matching (Hatch Act) grants to State Agricultural Experiment Stations located at the State Land Grant Universities set up by the Land Grant College Act of 1864; and 3) the Economic Research Service. The budgets for the three services for fiscal year 1985 were $470 million, $320 million, and $50 million, respectively.[17] In addition, the States provide more than $500 million a year in research funding to their agricultural experiment stations. The evolution of Federal and State support for research and extension directed at improvements in agricultural production technology is presented in table 3 below.[18]

Many econometric studies of the productivity return to agricultural research have been carried out in the past three decades, beginning with Zvi Griliches' classic 1958 study of hybrid corn technology. All but one of the studies have shown a very high internal rate of return on public sector agricultural research, as can be seen from table 4 below. The rate of return varies from a low of 21 percent to a high of 110 percent, with the vast majority in the 33 to 66 percent range. Public sector agricultural research has generally been considered to have been a significant success. Richard Nelson summarizes the characteristics of the agri-

"National Science Foundation, ***Federal R&D Funding by Budget Function, Fiscal Years 1984-1986,*** NSF 85-318 (Washington, DC: NSF, March 1985), pp. 74-75.

[18]R.E. Evenson, "Agriculture," ch. 5 of Richard R. Nelson ed.), *Government and Technical Progress: A Cross-Industry Analysis* (New York: Pergamon Press, 1982), p. 252.

Table 3.— Expenditures by the Public Sector on Research and Extension Oriented to Improved Agricultural Production Technology, 1890.1970

	Expenditures on research State agricultural experiment stations				USDA funded	Expenditures on Public
Year	Total[a]	% State funded	% Federally funded	% USDA funded	outside State[a]	Extension Service[a]
1890	97	22	78		2.6	0.3
1900	12.2	34	66		10.4	1 3
1910	370	39	61		474	2 4
1915	342	72	28		62.5	18,0
1920	287	77	23		49,0	464
1925	425	85	15		59.2	61 6
1930	75.6	73	27		96.5	772
1935	793	57	27	16	665	702
1940	113,2	54	28	18	1199	107. 7
? 945	1142	56	23	20	978	1020
? 950	1943	63	17	20	834	140.8
1955	251 4	63	17	20	89.2	152 1
1960	3448	55	15	30	876	169.5
1965	385.5	58	16	26	678	179.7
1970	414,5	66	16	18	109.5	2210
1975.	420.0	na	na	na	110,0	264.0
1980.	428.0	na	na	na	110.0	314.0

[a]In millions of constant 1980 dollars

SOURCE R E Evenson, "Agriculture," ch. 5 of Richard R Nelson (cd.), *Government and Technical Progress" A Cross-Industry Analysis* (New York Pergamon Press, 1982), p *252*

Table 4.—Econometric Studies of Productivity Return to Agricultural Research in the United States

Author (date)	Commodity	Time period	Rate of return
Griliches (1964)	Aggregate output	1949-59	35-40
Latimer (1964)	Aggregate output	1949-59	Not significant
Peterson (1967)	Poultry	1915-60	21
Evenson (1968)	Aggregate	1949-59	47
Cline (1975)	Aggregate	1939-48	41-50
Knutson and Tweeten (1979) , ,	Aggregate	1949-58	39-47
		1959-68	32-39
		1969-72	28-35
Bredahl and Peterson (1976)	Cash grain	1969	36
	Poultry	1969	37
	Dairy	1969	43
	Livestock	1969	47
Davis (1979)	Aggregate	1949-59	66-100
		1964-74	37
Evenson (1979)	Aggregate	1868-1926	65
		1927-50	95 (applied R&D)
		1927-50	110 (basic R&D)
		1948-71	45 (basic R&D)
Davis and Peterson (1981)	Aggregate	1949	100
		1954	79
		1959	66
		1964	37
		1969	37
		1974	37
Norton (1981)	Cash grain	1969	31[a]
	Poultry	1969	27
	Dairy	1969	56
	Livestock	1969	30
	Cash grain	1974	44
	Poultry	1974	33
	Dairy	1974	66

[a]Based on maximum lag length estimated (9 years)

SOURCE: Robert D Weaver, "Federal R&D and U S Agriculture An Assessment of the Role and Productivity Effects, ' paper presented at the National Academy of Sciences Workshop on "The Federal Role in Research and Development, " Nov 21-22 1985, p 27 cultural sector that have made it amenable to this success:

> In the first place, farming is an atomistic industry and farmers are not in competition with each other. Differential access to certain kinds of technological knowledge, or property rights in certain technologies, are not important to individual farmers. This fact at once means that farmers have little incentive to engage in R&D on their own behalf and opens the possibility that the farming community itself would provide the political constituency for public support of R&D.
>
> The Federal/State agricultural extension system . . . marshaled that support and put the farmers in a position of evaluating and influencing the publicly funded applied R&D. The system is highly decentralized. The regional nature of agricultural technology means that farmers in individual states see it to their advantage that their particular technologies be advanced as rapidly as possible.
>
> It was [the] combination of an evolving set of agricultural sciences based in the universities and supported publicly, and applied research and development also publicly funded but monitored politically by the farming community, that has made public support of agricultural technology as successful as it has been. Where private companies are funding significant amounts of innovative work and the industry is reasonably competitive, it is in the interest of the fanners as well as the companies that public R&D money be allocated to other things. [A] reasonably well defined division of labor has emerged between publicly and privately funded applied research.[19]

The nature of the agricultural sector explains why Federal R&D has a powerful effect and why econometric methods can arrive at relatively reliable estimates of this effect.

Aviation

Since World War II the Federal Government has provided a considerable amount of R&D support for aviation. Indeed, according to David Mowery, "the commercial aircraft industry is virtually unique among manufacturing industries in that a Federal research organization, the National Advisory Committee on Aeronautics (NACA, subsequently the National Aeronautics and Space Administration, NASA) has for many years conducted and funded research on airframe and propulsion technologies."[20] In addition, the Department of Defense has provided considerable support for research and development on military aviation that has generated considerable civilian spinoffs, and the aircraft industry itself conducts a great deal of in-house R&D. The total national R&D expenditure for aircraft from 1945 to 1982 was $104 billion (in 1972 dollars), of which $77 billion was provided by the military, $9 billion by NACA/NASA and $17.4 million by industry. Figure 1 breaks out those expenditures by year. About 45 percent of the total R&D budget went to aircrames, about 30 percent to avionics, and about 25 percent to engines.

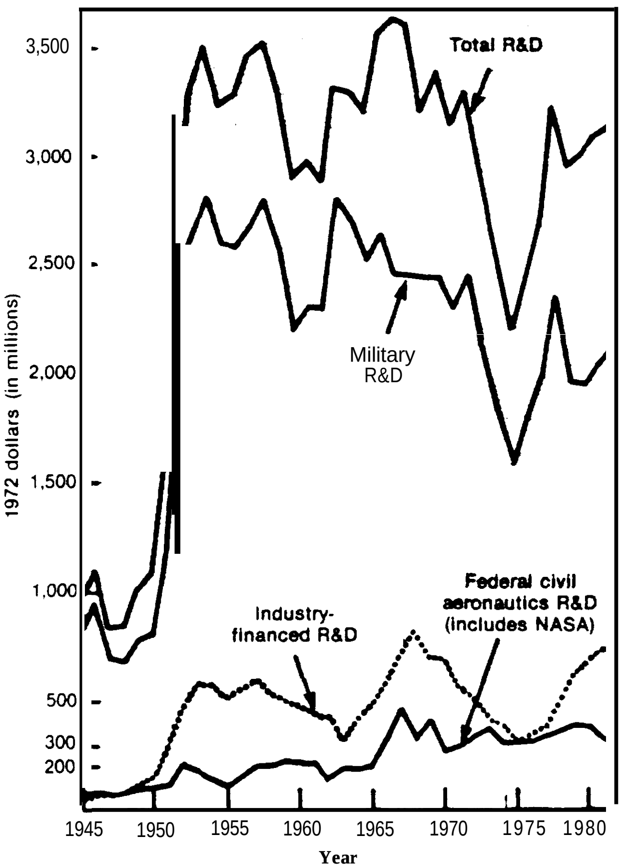
The benefits of this investment to the U.S. aviation industry and the consuming public were substantial. According to Mowery, the "total factor productivity in this [the commercial aviation] industry has grown more rapidly than in virtually any other U.S. industry during the postwar period. "[21] Two commonly used indices of aircraft performance are the number of available seats multiplied by the cruising speed (AS X Vc) and the direct operating costs per available seat mile (DOC). Between the DC-3 of 1940 and the Boeing 707 of 1959 the AS X Vc increased by a factor of 20 and the DOC fell by a factor of 3. The introduction of the Boeing 747 in 1970 increased the AS X Vc by another factor of 3 and halved direct operating costs. According to Mowery's calculations, if the total volume of airline passenger traffic in 1983 were to have been flown using 1939 technology (primarily the DC-3), the cost to the Nation would have been $24 billion (in 1972 dollars) rather than the $5.8 billion actually incurred (also in 1972 dollars). Thus, improvements in commercial aircraft technology led to more than $18.2 billion (in 1972 dollars) in additional air transportation services rendered for the actual amount paid. This benefit is considerably overstated in that consumers would have undoubtedly

[19] Richard R. Nelson (cd.), ***Government and Technical Progress: A Cross-Industry Analysis*** (New York: Pergamon Press, 1982), pp. 466-467.

[20] David C. Mowery, "Federal Funding of R&D in Transportation: The Case of Aviation, " paper presented to the National Academy of Sciences Workshop on 'The Federal Role in Research and Development, " Nov. 21-22, 1985, p. 13.

[21] Ibid., p. 6.

Figure 1.—Annual R&D Investment, 1945-82 (1972 dollars)



SOURCE: David C. Mowery, "Federal Funding of R&D in Transportation: The Case of Aviation, " paper presented to the National Academy of Sciences Workshop on "The Federal Role in Research and Development, " Nov. 21-22, 1985.

used other modes of transportation, or foregone a considerable amount of travel, had the aircraft operating costs not declined so substantially. On the other hand, the benefit does not take into account the value of the time saved by more rapid airline travel, the additional economic activity generated by an expanded airline industry, and the foreign trade benefits of the multibillion dollar sales of U.S. aircraft abroad.

If the $18 billion in additional air transportation services is taken as the benefit of improved aircraft technology, and the $104 billion total industry plus government R&D expenditure from 1945 to 1982 as the cost, then the social return from this investment appears to be on the order of 17 percent per year. If military development expenditures—about half of total military aviation R&D costs—am subtracted from the cost side, on the grounds that they do not directly support the commercial market, then the return on investment increases to close to 27 percent per year. Mowery has carried out a more sophisticated return on investment calculation in which he finds the internal rate of return from industry-financed and civilian Federal R&D to be about 24 percent.

Mowery also emphasizes that other factors play an important role. Civil Aeronautics Board regulations encouraged the adoption of new aircraft technologies between 1938 and 1978 by controlling air fares, which encouraged the airline industry to pursue a marketing strategy of technical innovation and service improvements. The regulatory incentive to innovate probably amplified the apparent social return.

Although it is not possible to isolate the civilian return on Federal aviation R&D, the dramatic expansion of the airline and aircraft industries in the United States after World War II is clear indication of the benefits of this unique Federal sectoral policy. Even in this industry, the economists are not unanimous on the productivity benefits. Terleckyj, for example, claims to have found "no effect of government R&D" on productivity improvements in the airline industry.[22] That his conclusions clash with common sense and other analyses illustrates the danger of depending solely on economic formulas to guide Federal R&D decisions.

Health

In 1985, the United States spent $381.2 billion on health care. Three percent of that, $11.5 billion, went to health R&D. The Federal Government funded two-thirds of the research, with half the money going to the National Institutes of Health (NIH).[23] With health care accounting for more than 10 percent of GNP and health research

[22]Terleckyj, "Measuring Economic Effects of Federal R&D Expenditures: Recent History With Special Emphasis on Federal R&D Performed in Industry, " op. cit., p. 6

[23]U.S. Department of Health and Human Services, *NIH Data Book*, MH Publication No. ***85-1261*** (Washington, DC: DHHS, June 1985).

claiming 33 percent of all nondefense Federal R&D funds,[24] health policymakers have asked whether it is possible to measure the economic benefits of biomedical research and development in terms of the primary output, improvements in health.

Measuring the productivity of health R&D is complicated by the value-laden issue of setting comparable economic values on well-being, illnesses, diseases, and life span. The economic costs of illness and disease, embodied in a "cost of illness" model, inform the national health agenda and the allocation of Federal health research budgets. An agency of the Public Health Service, the National Center for Health Statistics, has been directed by Congress to calculate annually the economic costs of illness and disease.

In response to an NIH initiative in the early 1970s, Selma Mushkin and her colleagues attempted to use a human capital methodology to quantify the economic costs of disease between 1900 and 1975.[25] They first calculated straightforward direct costs: expenditures on hospital care, physician's services, nursing home care, drugs, and medical appliances. The human capital model also includes indirect costs, specifically morbidity costs, which are losses incurred when illness results in absence from employment or a disability removes someone from the work force; and *mortality costs*, losses due to premature death. Mortality costs, in particular, embody the human capital approach in that they value one's life according to one's earnings, or according to the market value of one's skills. They estimate present value of future losses due to premature death, taking into account life expectancies for different sex, race, and age groups, varying labor force participation rates, and the discount rate.[26]

Mushkin attempted to estimate the contribution to the observed reduction in the national mortality rate of a number of factors, such as better public health measures, improved working conditions and nutrition, better personal health care and higher income, and biomedical research. She treated the mortality rate as a function of five independent variables—economic factors, societal factors, environmental factors, provider characteristics, and a measure of technical advances attributable to biomedical research and development. She then used regression analyses and other statistical techniques to determine the coefficients of each of the five variables in the equation predicting mortality rates. Indicators of technical advances due to research and development proved especially difficult to find, so Mushkin and her colleagues were forced to treat the technology variable as a residual; any reduction in mortality not attributable to the other variables in the function was attributed to biomedical research.

Based on this model, Mushkin found that biomedical research accounted for 20 to 30 percent of the reduction in mortality between 1930 and 1975. She estimated that each l-percent increase in biomedical research funds lowered mortality rates 0.05 percent. She also estimated that 39 percent of the reduction in days lost due to illness could be attributed to the results of biomedical research. Using human capital theory she estimated the value of each premature death averted at $76,000 and each work-year gained when illness is averted at $12,250. With these dollar values inserted for the reduced mortality and morbidity attributed to biomedical research, Mushkin found a return of $145 to $167 billion on a $30 billion investment, equivalent to an internal rate of return of 46 percent.

Critics have attacked the cost of illness model on economic and ethical grounds. Stephen Strickland argues that applying such estimates to decisions on public spending "carries an unacceptable implication that people should be protected, or saved, in proportion to their economic productivity and personal earnings. "[27] Such an approach devalues or dismisses the lives of many older citizens, children, women working in the home, or underemployed minority groups. In addition, a person's earnings may vary significantly over

---

[24] American Association for the Advancement of Science, ***Research & Development, FY* 1986 (Washington,** DC: AAAS, **1985), p. 27.**

[25] Selma Mushkin, ***Biomedical Research: Costs and Benefits*** (Cambridge, MA: Ballinger Publishing Co., 1979).

[26] For extensive illustrations of the human-capital approach, see, B.A. Weisbrod, "Costs and Benefits of Medical Research: A Case Study of Poliomyelitis, " *Journal* ***of Political Economy, vol. 7Q,*** May - June 1971, pp. **527-544.** Also R. Fein, ***Economics*** *of Mental* Illness (New York: Basic Books, 1958). Also, S.J. Mushkin, "Health as an Investment, " *Journal of Political* ***Economy, vol. 70,*** October 1962, pp. 129-159.

[27] Stephen P Strickland, ***Research and the*** *Health of Americans* (Lexington, MA: Lexington Books, 1978), p. 45

time, making extrapolations unrealistic. Furthermore, depending on the data and methodology used, calculations of the cost of a disease to the economy can produce a wide range of conclusions. Table 5 illustrates the divergent estimates of three groups analyzing the costs of the same diseases.

Many important costs of illness are also less quantifiable, such as the psychological costs for the patient and family. Costs can extend beyond the immediate period of illness and go unaccounted for if the disease is chronic. Some illnesses entail intergenerational costs, creating immeasurable long-term effects.

Measuring productivity gains in health care raises additional questions for economists. First, how does one assign a value to the health of individuals? Because almost all medical bills are paid by a third party-the government or an insurance company—health care is not as subject to market forces as other industries, and it is impossible to know the value of the care to the consumer. One indirect measure of the economic benefits of public health is increased worker productivity, but this gives no value to the health of the retired or those outside the work force. Distinguishing the inputs to improved health is as difficult as measuring the outputs; changes may be due to improved biomedical technology, nutrition, environmental conditions, exercise habits, more widely available or affordable health care, or a host of other factors.[28]

Physician and economist Jeffrey R. Harris summarizes the knotty conceptual and methodological problems inherent in measuring economic returns on investment in biomedical research:

1. In the biomedical sciences, the separate contributions of basic and applied research in biomedicine can be difficult to distinguish. . . .
2. The separate contributions of public and private investments in biomedical R&D are similarly difficult to distinguish. . . .
3. Not all biomedical innovations have arisen from biomedical research and development. . . .
4. The relative contributions of domestic and foreign investments in biomedical R&D will become an increasingly important issue.
5. Assignments of improvement in health to specific biomedical innovations is not always possible. . , . Observed improvements in health may have resulted from public health measures or changes in life style and the environment. . . .
6. The economic valuation of improvements in health raises important conceptual questions. In particular, innovations that prolong life generally result in increased economic transfers from younger, productive generations to older, less productive generations.

[28] Charles L. Vehorn, et al., "Measuring the Contribution of Biomedical Research to the Production of Health," Research Policy vol. 11, 1982, p. 4.; and Mushkin, "Health as an Investment," op. cit., p. 133.

Table 5.—Economic Costs of Disease: Differential Estimates

Disease	NIH estimates provided to House Appropriations Committee[a]	Estimates of special Commissions[b]	Estimates developed by methodology of Social Security Administration[c]
Arthritis	$4 billion	$13 billion	$3.5 billion
Asthma	$187 million		$855 million
Blindness	$5.2 billion		$2.2 billion
Cancer	$15 to $25 billion (range)		$17.4 billion
Diabetes		$5.3 billion	$3.5 billion
Digestive disease	$16.5 billion		$17.5 billion
Heart, lung, and blood disease	$58 billion		$12 billion
Epilepsy	$4.3 billion		$522 million
Influenza	$700 million		$4,1 billion
Mental illness	$40 billion	$36.8 billion	$13.9 billion

[a] Estimates offered in the testimony of NIH Institute directors or staff or provided in supplementary materials in the course of hearings on NIH appropriations for fiscal year 1977
[b] Estimates developed, respectively, by the National Commission on Arthritis and Related Musculoskeletal Diseases and the National Commission on Diabetes
[c] Estimates developed with the help of Barbara S Cooper, Office of Research and Statistics, Social Security Administration

SOURCE Stephen P Strickland, Research *and the Health of Americans* (Lexington, MA" Lexington Books, 1978), p 46

7. Although considerable attention has been devoted to valuing loss of life, the state of the art in gauging improvements in quality of life is far less advanced.
8. . . . assessments [based on gains in productivity] may substantially understate the public's willingness to pay for certain innovations. [29]

The three sectors analyzed above have similarities that make them amenable to quantitative analysis. Federal agricultural R&D is particularly suited to econometric analysis because the farmers themselves do almost no research; the government plays a significant role in promoting the application of its research; the government is a major customer, both directly and through farm and trade policies; and productivity improvements are easy to measure. The aviation industry is also a long-established, well-defined sector, dominated by Federal support. Aviation R&D is heavily weighted toward the development end and often incremental, making it easier to trace the returns on research. Despite the ease of tracking the products of the aviation industry and the extensive historical records, the case illustrates the difficulty of defining the scope of outputs to be included in an economic analysis; should these include only improved air transportation services, overall improved transportation, or all the indirect social and economic benefits of the airline industry? The health sector is less tractable, but has received significant attention because of the large sums of Federal money involved. Many of the great advances in heath care come serendipitously from basic research, making it difficult to trace the return on investment.

These analyses do not reveal the extent to which return on private or Federal R&D investment of an industry depends on variables such as R&D intensity, size, degree of concentration in a few large companies, whether the industry is emerging or stagnant, extent of technological competition within the industry, capital intensity}', market position, and government influences other than R&D.

"Jeffrey R. Hams, "Biomedical Research and Development: Measuring the Returns on Investment, " currently unpublished typescript of the National Academy of Sciences, National Academy of Engineers, and Institute of Medicine, November 1985, and for a further discussion of the basic-applied distinctions and the pathways between the two, see Julius H. Comroe and Robert Dripps, "Scientific Basis for the Support of Biomedical Science, " *Science*, vol. 192, Apr. 9, 1976, pp. 105-111.

SPINOFFS AND SPILLOVERS: NASA

The bulk of Federal research supports the internal missions of agencies like the Department of Defense (DOD) and NASA. While not aimed at commercial products, this research contributes indirectly to the development of commercial products or processes. These "spinoffs" and "spillovers" differ from the direct economic impacts of research sponsored by the Department of Agriculture, NIH and the civil aeronautics program of NASA in that they are unintended byproducts of activities carried out primarily to support non-economic mission agency goals, such as the exploration of space and national security. The economic impacts of the NASA R&D programs have been studied with special thoroughness and will therefore be described in some detail.

Three different approaches have been used to estimate the economic benefits of NASA's programs, according to former NASA chief economist, Henry Hertzfeld.[30] First are macroeconomic studies similar to those described for R&D investments in the economy as a whole. Second are macroeconomic analyses of the direct and indirect benefits from inventions and innovations resulting from the NASA R&D programs. Third are studies of the patents and licenses resulting from space R&D programs, used as a measure of the transfer of technology to the private sector.

Macroeconomic Studies

The first macroeconomic study, carried out by the Midwest Research Institute in 1971, estimated productivity changes in the national economy and

[30]Henry R. Hertzfeld, "Measuring the Economic Impact of Federal Research and Development Investments in Civilian Space Activities, " paper presented to the National Academy of Sciences Workshop on "The Federal Role m Research and Developer:, " Nov. 21-22, 1985.

then subtracted changes due to capital, labor, and such non-R&D residuals as demographics, education, length of workweek, and economies of scale. The discounted rate of return to NASA-sponsored R&D, calculated by doing a least-squares regression of the remaining residual, came to 33 percent, or a seven to one benefit/cost ratio. This study has several major liabilities. Assigning residual benefits to one factor, such as R&D, is inherently dangerous because we cannot be certain about the importance of other, unknown residual factors. The study did not look specifically at NASA R&D, but simply gave it credit for a proportionate share of the benefits of all R&D. The lifetime for benefits was arbitrarily set at 18 years .31

Chase Econometrics, in 1975, carried out a far more sophisticated analysis of the economic impact of NASA R&D, again using a residual calculation from a production function. Chase calculated a "cumulative 'productivity' return to NASA R&D of 14 to 1," which "translated into an annual discounted rate of return of 43 percent to NASA outlays. " However, when asked to replicate its methodology in 1980, using an "updated and longer time series, " Chase found that "productivity changes from NASA R&D spending proved not to be statistically different from zero. " Hertzfeld concludes that the revised Chase study "shows that *due to the theoretical and data problems with the macroeconomic model and data sets available, this approach to finding aggregate economic returns to R&D expenditure is difficult at best, and probably impossible.* "3[2]

Macroeconomic Studies

The two major macroeconomic studies of the economic benefits of NASA R&D programs used the "consumer surplus" approach to estimating the value to society of introducing a new product or reducing the cost of an existing product. The "consumer surplus" approach assumes that many people would be willing to pay more than the market price for a new or improved product, and uses supply and demand curves to estimate that "surplus. " Based on this approach, Mathematical, Inc., in 197.5, estimated the overall benefits to society from four NASA-stimulated technologies-gas turbine engines, integrated circuits, cryogenics, and "NASTRAN," an advanced computer program dealing with structural analysis. They found that "over a ten year period from 1975 to 1984 the four technologies could be expected to return a discounted total of $7 billion (in constant 1975 dollars) in benefits that were attributable to NASA's involvement in their developmerit. " This could be compared to a total NASA budget in fiscal year 1975 of $3.2 billion. (But about $30 billion over that 10-year period in 1975.)33

In 1977, another consulting firm, Mathtech, Inc., conducted a benefit/cost analysis of products adopted by the private sector as a result of NASA's formal technology transfer program. Mathtech only estimated the costs to the private sector "of further developing and transferring the innovations rather than the costs of the initial space R&D development of the technology. " The assumption here was that the initial developmental costs would have had to be incurred for the space program whether the technology was transferred or not. Unfortunately, this makes a calculation of the return on the NASA investment impossible. However, the benefit/cost ratios to the private sector were most impressive: 4:1 for the cardiac pacemaker, 41:1 for a laser cataract tool, 68:1 for a nickel-zinc battery, 340:1 for zinc-rich coatings, and 10:1 for a human tissue simulator.[34]

Both the Mathematical and the Mathtech studies are undermined by important weaknesses in the "consumer surplus" theory. The demand curves used to calculate the surpluses in that theory are inherently unreliable when applied to new technologies that have no well- formed demand function and to evolving technologies whose demand functions change over time. In addition, both studies fail to compare benefits to NASA development costs.

Patent Analysis

A third approach is to study what industry does with the licenses and patent waivers granted by NASA. Analyzing patent waivers, in which NASA allows a company to patent an invention devel -

"Ibid., p. 9.
"Ibid., pp. 11-12, emphasis added.

"Ibid., pp. ***18-19.***
"Ibid., pp. 19-20.

oped under contract, Hertzfeld has found that the commercialization rate (total commercialized inventions divided by total waivers) averaged 20.8 percent over the period 1961-75, with electrical machinery, communications equipment, and instruments accounting for over 69 percent of all commercialized specific waivers. Of the more than 197 NASA patents licensed to industry, Hertzfeld found that 54 were commercialized between 1959 and 1979. This was still a very small fraction (1. s percent) of the more than 3,500 patents owned by NASA at the time.[35]

"Ibid., pp. 24-26.

Finally, Hertzfeld points out that another economic benefit of the NASA space R&D program has been the creation of a multibillion dollar satellite communications industry and a tenfold reduction in the cost of satellite communications. However, Hertzfeld stresses that "economic returns are not the primary reason for space investmerits, " and therefore "no economic measure or calculation can, by definition, encompass the entirety of the return to space investment. "[36]

[36]Ibid., p. 4.

IMPLICATIONS FOR USING ECONOMIC INVESTMENT MODELS TO GUIDE FEDERAL R&D DECISIONMAKING

The studies described above present a discouraging picture for the use of economic returns as a valid measure of the value or desirability of Federal research Funding. Although strong positive returns to private sector research funding in general, and basic research funding in particular, have been indicated by macro-level econometric studies, no such positive returns have been shown for Federal research spending. Using econometric models to estimate the aggregate rate of return to all Federal research pushes the methodology beyond its limited capabilities.

The fundamental stumbling block to placing an economic value on Federal R&D is that *improving productivity or producing an economic return is not the primary justification for most Federal R&D programs.* The basic justification for Federal support of R&D is to encourage research that is socially desirable, high risk, or in the national interest but that is unlikely to be funded by the profit-driven private sector. The very concept of measuring the return to Federal investment in research in economic terms is therefore inherently flawed.

Economists who have studied this issue describe these flaws must vividly. At a fundamental level, Federal research is a public good which cannot be valued easily in economic terms. As Peter Reiss has expressed it:

> Typically the [activities of] the Federal government . . . produce things that have no market values that economists can even begin to mea;-ure. There is no market price, for example, for most health advances, and there is no conceivable evaluation for . . . public goods . . . like a strong national defense. These things are just not quantifiable .37

In many cases substantial economic benefits, such as NASA spinoffs and the computer and microelectronics technology spawned by DOD research, were secondary to the primary political and national security missions.

Frank Lichtenberg has found that Federal procurement has a far greater and more positive effect on private R&D expenditures than does Federal R&D. This finding is consistent with that of a number of other economists.[38] The direct influences of R&D support keep company with indirect but powerful Federal influences on private sector R&D through patent law, macroeconomic policies, tax incentives, trade policies, technology transfer, antitrust practices, and regulation.

[37]The National Academy of Sciences, Committee on Science, Engineering and Public Policy, ***The Federal Role in Research and Development,*** typescript transcript (Washington, DC: NAS, Nov 21, 1985), p. .s3.
[38]Nelson, op. cit., pp. 459-462, *471-472.*

In addition, there are fundamental flaws with the econometric methodology for measuring returns on investment when applied to Federal R&D. First, macroeconomic studies measure the aggregate return to the total expenditure on past R&D. They do not provide any information on the incremental return to the marginal expenditure on future R&D, which is the concern of the policymaker. Another major stumbling block in econometric analyses is that they measure inputs (R&D investments) and outputs (productivity changes) while ignoring the process that goes on in between. That process is the critical stage of turning laboratory research into a tangible return —innovation and commercialization. Research cannot result in product or process improvement unless each step in the move from idea to market is successful: advanced development, pilot studies, legal blessing of patents and licenses, production, and marketing. These intermediate commercialization steps are as important a factor in the move from R&D to productivity as is the R&D investment itself. However, the Federal Government has direct control only over its R&D investment, and very little influence over commercialization in the private sector.

According to Hertzfeld, the production function model used in most econometric analyses "assumes that a formal relationship between R&D expenditures and productivity exists" but "skips a number of steps in the process . . . creation of new knowledge, which will lead to ideas, inventions, innovations, and eventually, with proper marketing and distribution, commercial products . . . *Assuming that all of these missing steps take place from any given set of R&D expenditures is taking a giant leap without looking.* 39

Hertzfeld concludes his study of the economic impact of Federal R&D investments in civilian space activities with a statement that seems applicable to the more than 80 percent of the Federal research budget that is not aimed directly at improving economic competitiveness:

> . . *no economic study should attempt to put a "bottom line" ratio or return on space R&D investments.* There is no such number in existence—it only lives in the uncharted world of general equilibrium theory. . . . All such numbers are products of economic models with many limiting assumptions. Even when these assumptions and qualifications have been carefully laid out, the existence of the number is an attractive bait to those politicians and others who need to justify space R&D. Once a "total" returns number is used, it quickly finds its way into misuse .40

Clearly, R&D expenditures may be conceptualized as an investment flow, largely on the notion that R&D expenditures support the growth of an "R&D capital stock" or knowledge base, which contributes to economic growth and productivity improvement. However, a number of scholars recently have argued that the overenthusiastic application to investment decisionmaking of principles of portfolio management has led U.S. firms to underinvest in new technologies because of an "extreme of caution" (see ch. 4). Hence, even if satisfactory methodologies to calculate return could be developed, it might be unwise to adopt them.

R&D investment, and investment in basic research in particular, are characterized by high levels of *uncertainty* about their outcomes. A market that does not yet exist cannot be measured. While quantitative models of financial and other investment planning do provide methods for reducing *risk,* there exist virtually no models that can incorporate uncertainty. An explanation of this problem requires a brief description of the differences between risk and uncertainty.

Risk exists when decisionmakers can define the range of possible outcomes and assign probabilities to each. Models of risk analysis and reduction that have been developed for financial and investment decisionmaking rely heavily on the ability to quantify risk by assigning a probability to a specified set of likely, discrete outcomes.

---

[39] Hertzfeld, op. cit., p. 7.

*Ibid., p. 42, emphasis added.

[41] Spence's analysis of "Investment, Strategy and Growth in a New Market" explicitly disavows any attempt to deal with uncertainty:

> And finally, and less happily, uncertainty about demand, technology, the rates of entry, and competitors' behavior—all of which are practical problems for firms and inherent features of most new markets—is set aside to focus on the issue of the optimal penetration of the market and the dynamic aspects of strategic Interact Ion Integrating uncertainty into an appropriate model remains a high priority research topic

A.M. Spence, "Investment Strategy and Growth in a New Market," *Bell Journal of Economics,* vol. 10, No. 1, 1979, pp. 1-19 (quotation from p. 2).

True uncertainty, as defined by Frank H. Knight and others,[42] is the inability to specify possible outcomes. Estimating consequences, and therefore risks, is impossible. If one cannot specify outcomes, possibly due to the uniqueness of the process (e.g., the creation of an entirely new technology) or the lack of historical data on relations between actions and outcomes, quantitative models for risk-minimizing investment strategies are not applicable.

R&D investment, especially basic research investment, is a classic example of investment undertaken under conditions of severe uncertainty. Not only are research outcomes dimly perceived, assigning credible probabilities to the possible outcomes is impossible. Quantitative models developed to assess risk in energy exploration or financial management cannot address the uncertainty inherent in basic research spending decisions. As one moves from basic research to applied research or development, quantifiable risk replaces uncertainty. For this reason, quantitative models are likely to be far more useful in the evaluation of applied R&D or development decisions than for exploratory research. In addition, quantitative models are more applicable to decisions about distributing R&D investments among research installations or performers within a single discipline than allocating basic research funds across competing fields.

For the wealth of reasons presented above, it is clear that using economic returns to measure the value of specific or general Federal research expenditures is an inherently flawed approach. The only exceptions to this rule are certain Federal R&D programs whose specific goals are to improve the productivity y of particular industries or industrial processes.

‘Nelson and Winter describe the search behavior of a firm in an uncertain environment as follows:

> The areas surveyed by a decisionmaker inside the firm may well include identifiable “alternatives” that could be explored, but these are only dimly perceived and it may not be at all clear which will turn out to be best. The process of exploring perceived alternatives, or ●xogenous ●vents, may bring to light other alternatives not even contemplated in the original assessments. . It is clearly inappropriate to apply uncritically, in the analytical treatments of that process, formalisms that posit a sharply defined set of alternatives. .

Richard R. Nelson and Sidney G. Winter, An ***Evolutionary Theory of Economic Change*** (Cambridge, MA: Harvard University Press, 1982), pp. 171-172.

Chapter 3

Noneconomic Quantitative Measures: The “Output” of Science

chapter 3

Noneconomic Quantitative Measures: The "0utput" of Science

Having identified severe drawbacks to the use of econometric models to evaluate Federal R&D, OTA looked elsewhere for objective quantitative measures. The only quantitative approach to the evaluation of research output is bibliometrics, which analyzes scholarly publications for indications of quantity and quality. The underlying assumption of this approach is that knowledge is the essential product of research and publications are the most readily identified manifestations of that knowledge. With a gradually evolving methodology, bibliometricians have attempted to measure objectively the quantity and quality of research results. They have achieved some success in comparing projects within a discipline, and less in comparing disciplines. Bibliometric analysis does not address the most important policy question: how to compare the value of Federal research with other Federal programs.

A considerable amount of quantitative information about the U.S. science and engineering enterprise is published regularly by the National Science Foundation (NSF), the National Institutes of Health (NIH), and the National Research Council (NRC) in their reports on funding, personnel, degree attainment and graduate education. Every 2 years NSF publishes a 300-page compilation of this information, Science *Indicators.* Science indicators could be used to provide a rough measure of the health of the research enterprise in the United States if some agreement could be achieved in the science policy community about which of the thousands of numbers published by NSF are most relevant to that task. The use of science indicators to measure the quality of the research process in the United States is discussed later in this chapter.

BIB BIOMETRICS

The quantitative analysis of scientific publications is in its second generation. The first generation, spurred by Eugene Garfield's founding of the *Science Citation Index* and Derek de Solla Price's efforts, ' explored the feasibility of understanding science through its literature alone. Price boldly named this approach the "science of science" and published demonstrations of its heuristic, if not immediate policy, value, [z]

The second generation, now a decade old, sought to develop and exploit publication and citation data as a tool for informing decisionmakers, especially in Federal agencies and universities.[3] This current generation has many of the features 01 an

For a first person retrospective, see Eugene Garfield, ***Essays of an Information Scientist, vol.*** 1, **1962-73; vol. 2, 1974-76 (Philadelphia,** PA: ISI Press, 1977). For examples, see Eugene Garfield, et al., The ***Use of Citation*** **Data** ***in Writing the History of Science*** (Philadelphia, PA: Institute for Scientific Information, 1964); Derek de Solla Price, "Networks of Scientific Papers," Science, vol. 149, July 30, 1965, pp. **510-515; Derek de** Solla Price, "Is Technology Historically independent of Science? A Study in Statistical Historiography, " ***Technology and Culture, vol. 6,*** fall 1965, pp. **553-568.**

'Derek de Solla Price, *Little* ***Science, Big*** *Science* (New York Columbia University Press, 1963).

[3]For reviews, see Yehuda Elkana, et al. (eds.), *Toward A Metric of* Science: ***The Advent of Science Indicators (New*** York. John Wiley & Sons, 1978); Francis Narin, "Objectivity Versus Relevance in Studies of Scientific Advance, " *Scientometrics,* ***vol. 1,*** September 1978, pp. 35-41. The use of projected citation data in a controversial promotion and tenure case is described in N.L. Geller, et al., "Lifetime-Citation Rates to Compare Scientists Work, " Social ***Science Research,*** vol. ***7, 1978,*** pp. 345-305

institutionalized scientific specialty: multidisciplinary journals and practitioners, a clientele (both consumers and patrons), and numerous claims to the efficacy of "bibliometrics" as a policy tool.[4] The quantitative analysis of scientific publications has arguably established its place in the evaluation of research outcomes and as an input both to the allocation of resources for research and to the expectation that the growth of scientific knowledge can be measured, interpreted, and indeed, manipulated.

This chapter focuses on the second generation of noneconomic quantitative measures of scientific research results and evaluates its usefulness to policymakers. The chapter assesses the most promising approaches and methods that have been employed and suggests how quantitative data and models could be refined to augment decisionmaking processes in science.

The First Generation of Bibliometrics (1961-74)

The pioneers of bibliometrics searched for ways to understand science independent of the scientists themselves. First-person accounts, questionnaires, and historical narratives all require some form of cooperation or consent of the scientists involved. This dependence on self-interest sources could bias the results. Bibliometric pioneers of the early 1960s saw a need first to reconstruct, then to monitor and predict, the structure and products of science. Eugene Garfield and Derek de Solla Price talked about "invisible colleges" and the tracing of "intellectual influence" as a mirror held up to science, imperfect but public, using the formal communication system of science. Science literature could be studied—without recourse to the authors—to open new vistas, both practical and analytical, once it was cataloged, indexed and made retrievable.

With the creation of the Science *Citation Index* (SCI), the scientific literature became a data source for the quantitative analysis of science. It generated both the concepts and measurement techniques that formed the bedrock of bibliometrics.[5] These include the principal units of analysis: publications (papers, articles, journals), citations (bibliographic references), and their producers (individual authors and collaborators in teams). When subjected to the primary methods of analysis—counting, linking, and mapping—these units yield measures of higher order concepts: coherent social groups, theory groups, networks, clusters, problem domains, specialties, subfields, and fields.

Computers aided the increasingly sophisticated manipulation of documents in the growing SCI database. Journal publications could be counted by author, but also aggregated into schools of

[4]Cofounded in *1978* by Garfield and Price, *Scientometrics* became the flagship journal of bibliometrics. Its contributors seem to come primarily from information science, psychology, and sociology. Other spurs to the institutionalization and visibility of bibliometrics has been, since **1972, the National** Science Board's biennial ***Science*** Indicators series and the ongoing work of the Institute for Scientific Information (especially Henry Small) and Francis Narms Computer Horizons, Inc. (discussed below). For historical perspectives on the development of this specialty, see Daryl E. Chubin, "Beyond Invisible Colleges: Inspirations and Aspirations of Post-1972 Social Studies of Science, " *Scientometrics, vol. 6, 1985,* pp. 221-254; Daryl E. Chubin and S. Restive, "The 'Mooting' of Science Studies: Strong Programs and Science Policy," in K.D. Knorr-Cetina and M. Mulkay (eds.), Science Observed (London and Beverly Hills, CA: Sage, 1983), pp. *58-83. Also see* the special issue of *Scientometrics, vol. 6, 1985,* dedicated to the memory of Derek Price.

[5]These are touted, debated, and assailed in Daryl E. Chubin, "The Conceptualization of Scientific Specialties, ' *The Sociological Quarterly, vol. 17,* autumn 1976, pp. *448-476;* Daryl E. Chubin, "Constructing and Reconstructing Scientific Reality: A Meta-Analysis, " *International Society for the Sociology of Knowledge Newsletter, vol. 7,* May **1981,** pp. 22-28; Susan E. Cozzens, "Taking the Measure of Science: A Review of Citation Theories, " *ISSK Newsletter, vol. 7,* May **1981,** pp. **16-21**; D. Edge, "Quantitative Measures of Communication in Science: A Critical Review, " *History of Science vol. 17, 1979,* pp. *102-134;* and in various chapters in Elkana, et al., op. cit.

thought or whole institutions.[6] Consistent and influential contributors to the literature could be identified by co-citations (the number of times two papers are cited in the same article) and separated from occasional authors. The resultant co-citation clusters could be depicted as a "map of science" for a given year showing the strength of links within clusters and the relations, if any, among them.[7]

By the mid-1970s, bibliometricians were constructing structural and graphical maps of the domains and levels of research activity in science. Further, they were comparing these pictures to other accounts, built on biographic and demographic information, informal communication, and other informant-centered data, to depict how research communities—their research foci, intellectual leaders, and specialized journals-change over time. They thus offered a more comprehensive perspective on the growth of knowledge, at least in terms of its outputs, than was ever previously available. " Analysts differed in their interpretation and application of the data, and the life of the community responsible for the outputs tended to remain unobserved. Nevertheless, bibliometric analysis began to offer the promise of the independent baseline implied in Price's phrase "the science of science. "[9]

The Second Generation (1975=85)

The legacy of the first generation was the promise of its scholarly literature. The second generation has attempted to deliver on the promise that bibliometric analysis could be predictive and reliable for decisionmaking. That promise has yet to be fulfilled, for reasons that will be discussed below. However, there is growing evidence that the quantitative assessment of science warrants the attention it is now receiving from policymakers both in the United States and Europe.

The analysts discussed below have used bibliometrics to anticipate the source of "greatest contributions" and identify promising research projects. They produce policy-relevant documents and recognize intervention decisions as a desirable consequence of their work. Several governments have funded their efforts. A look at the leading

[6]Seminal work here is D. Crane, ***Invisible Colleges: Diffusion of Knowledge in Scientific Communities*** (Chicago and London: University of Chicago Press, 1972); B.C. Griffith and 14. C. Mullins, "Coherent Social Groups in Scientific Change, " *Science* ***vol. 177***, Sept. 15, 1972, pp. ***959-964;*** N.C. Mullins, 'The Development of a Scientific Specialty: The Phage Group and the Origins of Molecular Biology, " *Minerva*, ***vol. 10, 1972,*** pp. 52-82; N.C. Mullins, ***Theory and Theory Groups in Contemporary American Sociology (New*** York: Harper Row, ***1973);*** **and Garfield, op. cit., "Corporate** Index" **that** lists publications by institution of author.

'The methodological groundwork for co-citation analysis is presented in B.C. Griffith, et al., "The Structure of Scientific Literatures II: Toward a Macro- and Microstructure for Science, " ***Science Studies,*** vol. 4, 1974, pp. 339-365; H.G. Small, "Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents, " *Journal of the* ***American Society for Information Science,*** vol. 24, 1973, pp. 265-269; H.G. Small, "Multiple Citation Patterns in Scientific Literature: The Circle and Hill Models, " ***Information Storage and Retrieval, vol.*** 10, 1974, pp. 393-402; H.G. Small and B.C. Griffith, 'The Structure of Scientific Literatures I: Identifying and Graphing Specialties," Science ***Studies, vol.*** 4, 1974, pp. 1740.

'Noteworthy illustrations are discussed in Daryl E. Chubin, 'The Conceptualization of Scientific Specialties, " op. cit., and G.N. Gilbert, "Measuring the Growth of Science: A Review of Indicators of Scientific Growth, " *Scientometrics,* vol. 1, September 1978, pp. 9-34.

'For example, in 1969, Price's "Measuring the Size of Science, " ***Proceedings of the Israel Academy of Sciences and*** *Humanities*, vol. 4, 1969, pp. 98-111, tied national publication activity to percent of GNP allotted to R&D. By 1975, F. Narin and M. Carpenter ("National Publication and Citation Comparisons, " *JASIS*. voi, 26, pp. 80-93) were computing shares, on a nation-bv-nation basis, of the world literature, and characterizing interrelations among journals (Francis Narin, et al., "Interrelationship of Scientific Journals,' *JASIS* vol. ***23, 1972,*** pp. ***323-331), as*** well as the content of the literature in broad fields (Francis Narin, et al., "Structure of the Biomedical Literature, " *JASIS*, vol. ***27, 1976,*** pp 25-45). These analyses employed algorithms for tallying, weighing, and linking keywords in article titles to citations aggregated to journals and authors nation -of-affiliation at the time of publication. Some would call this methodology "crude"; others would herald its sophistication for discerning patterns in an otherwise massive and perplexing literature The latter is preciseiy the mentality guiding the *Science Indicators* volumes and foreshadowed in two other pioneering papers of the first generation: Eugene Garfield, "Citation Indexing tor Studying Science, " ***Nature***, vol. ***227, 1970,*** pp. ***659-671;*** Derek de Solla Price, "Citation Measures of Hard Science, Soft Science, Technology and Non-science, " ***Communication Among Scientists and*** *Engineers*, ***C.*** Nelson and D. Pollock (eds.) (Lexington, MA: D.C. Heath, 1970), pp. 3-22.

practitioners will reveal how they approach the question:

> How can we characterize the effects of decisions about funding programs as they reverberate into the various levels of the scientific community: up from "fields" into disciplines and down from "fields" into research areas or teams?[10]

Francis Narin of Computer Horizons, Inc., is the veteran performer, linking the two generations. His computerized approach is based on the components of the Science *Citation Index* and used in conjunction with other data, such as the National Library of Medicine's MEDLINE and the NIH in-house grant profile system, Information for Management Planning, Analysis, and Coordination (IMPAC). Although Narin's work tends toward the macroscopic, its manipulations have grown more sophisticated in their capability of addressing micro-level questions. Narin's methodology answers quantitatively the following kinds of questions:

- Are articles published in basic journals referenced in clinical and practitioner journals [these types derive from Narin's own classification of article content in journals]?
- Is there a relationship between priority scores on research applications and number of articles produced and citations received?
- Are grants to medical schools more productive than grants to academic departments?
- Are young researchers more productive than older researchers?
- Is the return on investment mechanism [investigator-initiated proposals] more productive than other support mechanisms?
- How often do National Institute on Drug Abuse, National Institute on Alcohol Abuse and Alcoholism, and other NIH-supported researchers cite work supported by the National Institute of Mental Health?

A common criticism of Narin's work is that it is too descriptive and relies on ad hoc explanation for the observed patterns and trends. Some feel it is excessively dependent on a literature baseline and does not reflect an understanding of the sciences it appraises. In a current project sponsored by the National Cancer Institute (NCI), "An Assessment of the Factors Affecting Critical Cancer Research Findings," Narin consciously tries to remedy the problem by working closely from the outset with a panel of cancer researchers. He is tracing key events through participant consensus, the historical record, and various bibliometric indicators. Discrepancies are apparently negotiated as the project unfolds, though the exact negotiation procedure is not specified .11

Another departure for Narin stems from his acquisition and computerization of U.S. Patent Office case files that will permit mapping of literature citations in patents at the national, industry, and inventor levels. An infant literature has crystallized around the notion of "technology indicators" with patents signifying the conversion of knowledge into an innovation with commercial and social value—another tangible return on investment.[12]

Irvine and Martin's (Science Policy Research Unit, University of Sussex, UK) evaluation program in "converging partial indicators" has gained attention for three important reasons:

1. They claim to assess the basic research performance of large technologydependent facilities, such as the European Organizations for Nuclear Research (CERN) accelerator and the Isaac Newton Telescope.
2. They have made cost-effectiveness the central performance criterion in their input-output scheme.
3. Their "triangulation" methodology is an impressive codification of many separate proce-

[10]Susan E. Cozzens, "Editor's Introduction," m "Funding and Knowledge Growth," Theme Section, *Social Studies of Science*, vol. 16, February 1986, forthcoming (quote from mimeo version, p. 9).

[11]The objective of this project is to estimate knowledge returns from the U.S. war on cancer. What has been the extent and character of NCI funding in the cancer literature: are highly cited papers and authors supported by NCI grants and contracts? More on this genre of study is presented below and in other chapters of this technical memorandum, but see Francis Narin and R.T. Shapiro, "The Extramural Role of the NIH as a Research Support Agency," *Federation Proceeding*, vol. 36, October 1977, pp. 2470-2475.

[12]M.P. Carpenter, et al., "Citation Rates to Technologically Important Patents," *World Patent Information, vol. 3, 1981*, pp. 161-163; and various case study reports on patent activity emanating from Battelle's Pacific Northwest Laboratories, for example, R.S. Campbell and L.O. Levine, *Technology indicators Based on Citation Data: Three Case Studies*, Phase II Report prepared for the National Science Foundation Grant, PRA 78-20321 and Contract 2311103578, May 1984.

dures and measures that have been advocated both by policymakers and analysts. '3

The synopsis presented below is based primarily on a review of Irvine and Martin's articles, four critiques, and a reply.[14] The Irvine and Martin rationale for developing indicators of past research performance is to provide "a means to keep the peer-review system 'honest.' " Irvine and Martin caution us further "to distinguish between conventional peer-review (involving a small number of referees or 'experts' on a panel) and our extensive peer-evaluations drawing in very large numbers of researchers across different countries and based on structured confidential interviews and attitude surveys."[15] For the two investigators, conventional grants or journal peer review is but a single indicator; when combined with bibliometric data on research performance and external assessments of the likely future performance of new facilities, a series of multiple indicators is formed. If the indicators converge, Irvine and Martin regard the evaluation results as relatively reliable.[16]

As proxies, partial indicators must stand for a lot that goes unmeasured—by choice or otherwise. Sometimes the interpretive burden is overwhelming (see table 6). No matter how systematic, quantitative, and convergent their findings appear, Irvine and Martin's use of triangulation is problematic, as they admit (see table 7 for a summary):

> The fact that the indicators converge in a given case does not "prove" that the results are 100 percent certain—the indicators may all be "wrong" together. However, if a research facility like the Lick 3-meter telescope produces a comparatively large publication output at fairly low cost, if those papers are relatively highly cited, . . . and if large numbers of astronomers rate it highly in the course of structured interviews, we would place more credibility on the resulting conclusion that this was a successful facility than if the same finding were arrived at by a panel of three or four "experts" without access to the systematic information that we have collected.[17]

If the output measures do not converge, the results become quite problematic. There is no straightforward means of resolution except intuition and judgment.

Tables 7, 8, and 9 present typical samples of the information one can obtain from Irvine and Martin's analyses. Table 7 shows for four different but comparable optical telescopes the average number of papers published per year over the decade 1969-78, the cost per paper, the number of citations to work done on that telescope over the 4-year period 1974-78, the average number of citations per paper, and the number of papers cited 12 or more times. This table was part of a paper that demonstrated that the Isaac Newton Telescope (INT) in Great Britain was more costly and less productive than several comparable facilities. (This was largely due to a political decision to locate the INT at a poor observing site on British soil. Subsequently, it was moved to a more favorable site at La Palma.) The table compares the various facilities in terms of output (papers per year), cost-effectiveness (cost per paper), influence (citations), and significance of scientific work (citations per paper and number of papers cited more than 12 times).

Table 8 presents similar output data for world experimental high energy physics facilities from 1977 to 1980. The table shows, for example, that although the largest number of papers were produced at the CERN proton synchrotrons in 1978, this facility did not have the greatest influence in terms of the number of citations to work done there, nor was it producing the most significant

[13] J. Irvine and B.R. Martin, Foresight *in Science: Picking the Winners* (London: Frances Pinter, 1984). See especially B.R. Martin and J. Irvine, "Assessing Basic Research: Some Partial Indicators of Scientific Progress m Radio Astronomy," *Research Policy,* vol. 12, 1983, **pp. 61-90.**

"The five components are: J. Knge and D. Pestre, "A Critique of Irvine and Martin's Methodology for Evaluating Big Science," *Social Studies of* Science, vol. 15, 1985, pp. **425-539;** H.F. Moed and A.F.J. van ***Ram, "Critical*** Remarks on Irvine and Martin's Methodology for Evaluating Scientific Performance," ***Social Studies of Science, vol.*** 15, 1985, pp. **539-547; R. Bud,** 'The Case of the Disappearing **Caveat: A Critique** of Irvine and Martin's Methodology," ***Social Studies of Science, vol.*** **15, 1985,** pp. 548-553; H.M. Collins, 'The Possibilities of Science Policy," ***Social Studies of Science,*** VOI. 15, 1985, pp. **554-558; and** B.R. Martin and J. Irvine, "Evaluating the Evaluators: A Reply to Our Critics," ***Social Studies of Science, vol.*** **15, 1985, pp. 558-575.** For brevity, quotes from the critics will be noted in the text by (page number) only, those from Martin and Irvine as (IM, page number).

'Ibid., p. 566.

"Ibid., p. 527.

'"Ibid., p. **568**

Table 6.—Main Problems With the Various Partial Indicators of Scientific Progress and Details of How Their Effects May Be Minimized

Partial indicator based on	Problem	How effects may be minimized
A. Publication counts	1. Each publication does not make an equal contribution to scientific knowledge	Use citations to indicate average impact of a group's publications, and to identify very highly cited papers
	2. Variation of publication rates with specialty and institutional context	Choose matched groups producing similar types of papers within a single specialty
B. Citation analysis	1. Technical limitations with *Science Citation Index:*	
	a. first-author only listed	Not a problem for research groups
	b. variations in names	
	c. authors with identical names	Check manually
	d. clerical errors	
	e. incomplete coverage of journals	Not a serious problem for "Big Science"
	2. Variation of citation rate during lifetime of a paper—unrecognized advances on the one hand, and integration of basic ideas on the other	Not a problem if citations are regarded as an Indlcator of impact, rather than quality or importance
	3. Critical citations	
	4. "Halo effect" citations	
	5. Variation of citation rate with types of paper and specialty	Choose matched groups producing similar types of papers within a single specialty
	6. Self-citation and "in-house" citation (SC and IHC)	Check empirically and adjust results if the incidence of SC or IHC varies between groups
C. Peer	1. Perceived implication of results for own center and competitors may affect evaluation	1. Use a complete sample, or a large representative sample (25°/0 or more)
	2. Individuals evaluate scientific contributions in relation to their own (very different) cognitive and social locations.	2. Use verbal rather than written survey so can press evaluator if a divergence between expressed opinions and actual views is suspected
	3. "Conformist" assessments (e.g., "halo effect") accentuated by lack of knowledge of contributions of different centers	3. Assure evaluators of confidentiality
		4. Check for systematic variations between different groups of evaluators

SOURCE: B.R. Martin and J. Irvine, "Assessing Basic Research: Some Partial indicators of Scientific Progress in Radio Astronomy," Research *Policy*, vol 12, 1983

Table 7.—Output indicators for Optical Telescopes—A Summary

	Lick 3-meter	KPNO[a] Z. l-meter	CTIO[b] 1 ,5-meter	INT[c] 2,5-meter
Average number of papers pa., 1969-78	42	43	35	7
Cost per paper m 1978	f13k	f7k	f6k	f63k
Citations to work of past 4 years m 1978 . . .	920	710	580	140
Average citations per paper in 1978, ., .,	4.2	3.3	3.3	3.6
Number of papers cited 12 or more times in a year, 1969-78, .	41	31	21	4

[a]Kitt Peak National Observatory (U.S.).
[b]Cerro Tololo InterAmerican Observatory (Chile).
[c]Isaac Newton Telescope (Great Britain).

SOURCE: B.R. Martin and J. Irvine, "Evaluating the Evaluators: A Reply to Our Critics," *Social Studies of Science*, vol. 15, 1985, p. 569.

scientific work in terms of average citations per paper or number of highly cited papers. One can also see the decreasing importance of the CERN proton synchrotrons as newer machines such as the CERN super proton synchrotrons and the German Electron Synchrotrons Laboratories (DESY) accelerator at Hamburg come on-line and begin to produce important results.

Table 9 presents the high energy physicists' own evaluations of the relative contributions of the different facilities described in table 8, based on a mail survey of 182 researchers in 11 countries. These evaluations are based on the relative outputs of the different accelerators over their entire

Table 8.—Experimental High-Energy Physics, 1977-80

	Percent of papers published in past 2 years		Percent of citations to work of past 4 years		Average citations per paper		Highly cited papers: number cited in times			
	1978	1980	1978	1980	1978	1980	n >15	n >30	n >50	n >100
CERN proton synchroton	22.0°/0	11.5 %	14.50/0	12.50/o	2.2	2.2	13	2	1	0
Brookhaven/AGS	5.50/0	5.5%	5.0 "/0	3.0%	2.7	1.6	0	0	0	0
Serpukhov	12.0°/0	14.00/0	4.0%	5.0 0/0	1.2	1.2	0	0	0	0
CERN ISR[a]	4.5%	5.5 %	7.0%	7.5%	5.4	4.4	11	2	0	0
Fermilab	16.5°/0	1 9.0%	32.0%	21.5%	7.3	3.6	40	10	5	1
CERN super proton synchroton	2.50/o	8.50/o	4.0%	8.50/o	12.7	5.0	19	7	3	0
SLAC [b]	9.5%	6.00/0	15.0%	11.5%	5,7	4.4	26	6	1	1
DESY	4.0 "/0	6.50/o	5.5%	15.50/0	5.7	8.8	36	16	4	0
Rest of world	23.5°/0	24.00/o	13.0 "/0	15.00/0	2.0	1.9	19	5	0	0
World total	1,115	930	8,190	5,090	3.5	3.0	164	48	14	2
	100'%0	100%	100%	100%						

[a]Intersecting storage rings.
[b]Stanford Linear Accelerator Center.

SOURCE: J Irvine and B R Martin, "Quantitative Science Policy Research, " testimony to the House Committee on Science and Technology Oct 30 1985

Table 9.—Assessments (on a 10-point scale[a]) of Main Proton Accelerators in Terms of "Discoveries" and "Providing More Precise Measurements"

	Self -rankings	Peer-rankings	Overall rankings (sample size= 169)
Discoveries:			
Brookhaven/AGS	9.5(± 0. 1)	9.0(± 0.1)	9.2(± 0. 1)
CERN PS	7 1(± 0.2)	6.7(± 0.2)	6.9(± 0. 1)
CERN ISR	6.8(± 0. 3)	5.9(± 3.2)	6.1(±0.2)
CERN SPS	5.9(±0.3)	5.6(±0.2)	5.7(±0. 1)
Fermilab	7 4(± 0.3)	7.1 (±0. 1)	7.2(±0.1)
Serpukhov	3.8(±0.5)	2.6(±0.1)	2.7(±0. 1)
More precise measurements:			
Brookhaven/AGS	7 1(±0.2)	7,2(±0.2)	7 2(*o. 1)
CERN PS	8.5(±0. 1)	8.5(±0.l)	8.5(±0.. 1)
CERN ISR	7 3(*0.3)	6.9(±0..2)	7 0(±0. 1)
CERN SPS	8 2(±0.2)	8.2(*0.2)	8.2(±0. 1)
Fermilab	6.3(*0.2)	6.0(±0. 2)	6. 1(±0..1)
Serpukhov	4.3(±0.5)	3.5(*0.2)	3.6(±0..2)

[a]10-top. The assessements are based on the relative outputs from the accelerators over their entire operational careers up to the time of the !nterviews with hlgh-energy physicists in late 1981 to early 1982.

SOURCE J Irvine and B.R Martin, "Quantitative Science Policy Research, ' testimony to the House Committee on Science and Technology, Oct 30 1985

operational careers and therefore do not necessarily match the output indicators from table 8, which are for a 3-year period. Comparable indicators exist for the entire 22-year period, 1960-82, in Irvine and Martin's papers.

Overall, the contribution of Irvine and Martin's work to research evaluation can be summarized as follows:

- They have collected, synthesized, and published a colossal amount of information—all original data—about the scientific performance of big and expensive scientific institutes.
- They have shown that when peers are assessing their own fields they can be reliable judges of scientific performance.
- Where choices have to be made in a field, among several similar research units competing for resources, Irvine and Martin provide policymakers with sound information for assisting a rational decision.

On the negative side, it is not known how the Irvine and Martin approach would fare in non-Big Science areas. Would the methodology transfer, as Irvine and Martin assert, to different cultural and research contexts? Even if converging indicators can validate contribution to scientific progress or the impact of a research team on its peer community, judgments of applicability and quality of these findings do not automatically follow. These are properties of interpretation, not analysis. Irvine and Martin tend to confuse the two.

The Irvine and Martin methodology is based solely on bibliometric and peer ratings among facilities in the same science, not knowledge-producing facilities in different sciences. However, the strategic choices between fields are the tough ones in a zero-sum world. Like peer review, "converging partial indicators" are useless for strategic choices.

Finally, knowledge is produced by scientific communities, not individual institutions. Therefore, comparing facilities may bean empty exercise. The implication of Irvine and Martin's recommendations is that reducing or eliminating funding to the least cost-effective facility has no adverse effect on progress, and in fact diverts scarce resources to more productive facilities elsewhere. Such a strategy, however, undermines the knowledge-producing community and runs counter to the view of science as a cultural activity that intertwines local teams with distant peers through literature, informal communication, and the training of new generations of practitioners. These are not ignored by Irvine and Martin, but they are minimized.

The Center for Research Planning (Coward, Franklin, and Simon) has developed the method of "bibliometric modeling" based on Institute for Scientific Information (ISI) data and co-citation clusters to monitor the research front of a given specialty. [18] Each model consists of an intellectual base and the current work of the specialty. When brought together in a computer, these two sets of papers contain the building blocks of a model: "title words (keywords) of its current papers" and "the demographics of the specialty" (performing organizations and countries). The "age of its intellectual base" is an indicator of the specialty's "development potential. "

Working closely with the Economic and Social Research Council of the United Kingdom, the Center for Research Planning (CRP) built specialty-specific models and met in workshops with key participating research teams and technical experts from the respective research councils responsible for the funding. This hands-on approach allowed data to be passed to the scientists for their own independent analysis. In other words, CRP works with representatives of the knowledge-producing communities being evaluated and the policy users themselves to increase credibility and relevance of their studies. As their 1984 final report to the Advisory Board for the Research Councils states:

> The models are not intended to function as computer-based decision algorithms in the science resource allocation process, but should be viewed as a potential decision-support system .19

Though the CRP approach is, like Narin's, data-intensive and unobtrusive at its source, it is more interactive with the relevant actors. [20] It is unclear how this iterative and interactive process affects interpretations. While the modeling notion is made explicit by CRP, their methodology is not as well codified as Irvine and Martin's. Perhaps recognizing this, CRP is championing interactive computing with their models to moderate the suspicions of researchers and policymakers alike. Such interactions will allow users to ask specific questions of a model on-line and receive immediate answers. Though this innovation will have obvious appeal, the jury is still out on its efficacy.

Other Important Teams

In Holland, the research team of H.F. Moed, W.J.M. Burger, J.G. Frankfort, and A.F.J. Van Raam has carried out extensive comparative studies of the research productivity of different departments at their home University of Leiden. They have used publication and citation counts to track trends in the quantity and impact of research published by individuals and teams in the Faculties of Medicine and Mathematics and Natural Sciences over a 10-year period (1970-80).[21]

[18] H. R. Coward, et al., "ABRC Science Policy Study: Co-Citation Bibliometric Models" (abridged), presented to the Advisory Board for Research Councils, Department of Education and Science, United Kingdom, July 1984, pp. 1-3, 65.

[19] Ibid.

[20] L. Simon, et al., "A Bibliometric Evaluation of the U. S.-Italy Cooperative Scientific Research Program, " *Evaluation of U.S.-Italy Bilateral Science Program* (Washington, DC: National Science Foundation, February 1985); J.J. Franklin and H.R. Coward, "Planning International Cooperation in Resource-Intensive Science: Some Applications of Bibliometric Model Data, " papers presented at the National Science Foundation symposium entitled "International Cooperation in Big Science, " February 1985.

[21] H.F. Moed, et al., *On the Measurement of Research Performance: The Use of Bibliometric Indicators* (Leiden, the Netherlands: University of Leiden, Research Policy Unit, Diensten OWZ/PISA, 1983); H.F. Moed, et al., "A Comparative Study of Bibliometric Past Performance Analysis and Peer Judgment, " *Scientometrics*, vol. 8, Nos. 3-4, 1985, pp. 149-159; and H.F. Moed, et al., The Application of Bibliometric Indicators: Important Field- and Time-Dependent Factors to be Considered, " vol. 8, Nos. 3-4, 1985, pp. 177-203.

In France, the team of Michel Calon, Jean-Pierre Courtial, William Turner, and Ghislaine Chartron at the School of Mines in Paris has used the technique of "co-word" analysis to identify principal problem areas being worked on by the laboratones of a major French research institute and to situate that research in its international context. Co-word analysis monitors the number of times that keywords, identified by researchers as describing a research problem, occur in pairs in the research literature. A map of the pairings of these "co-words" can give one a sense of the structure of a research field .22

[22]A. Rip and M. Courtial, "Co-word Maps of Biotechnologies: An Example of Cognitive Scientometrics," *Scientometrics*, vol. 6, 1984, pp. 381-400. M. Callon, et al., "The Transition Model and Its Exploration Through Co-Word Analysis: Using Graphs for Negotiating Research Policies," Centre de Sociologie, Ecole des Mines de Pans and Centre Nationale de la Recherche Scientitique, mimeo.

THE USE OF BIBLIOMETRICS TO EVALUATE RESEARCH AT THE NATIONAL INSTITUTES OF HEALTH

Of all the Federal agencies supporting R&D, NIH conducts the most extensive ex post evaluation of its research through bibliometric studies and other activities carried out at the individual institutes. In 1970, the Public Health Service Act was amended to set aside for evaluation activities up to 1 percent of the funds appropriated to any program authorized by the Act for evaluation. Each of the 11 institutes of NIH receives a separate appropriation from Congress, so that each can evaluate its own programs. The Program Evaluation Branch in the Office of the Director studies cross-cutting issues, develops new approaches to evaluation, and supports the development of data resources for this purpose .23 The budget for evaluation studies at NIH was $5.8 million in 1985 ($2.8 million from set-aside funds and $3 million from the regular budget). A review of NIH's use of bibliometrics illustrate the range of useful information that can be produced.

NIH Databases

NIH maintains several extensive databases that are used for evaluation. The IMPAC database contains detailed information about all active reviews and awards of NIH grants, including the names of all principal investigators, the applicant's institution, the type of grant, the review group, priority score awarded through peer review, the funding institute, and the amount of support. Two longitudinal databases developed from IMPAC track the training or funding history for any investigator who has applied for NIH support. A separate financial database holds year-end data on all appropriations and obligations since 1950 for all NIH institutes and mechanisms of support .[24]

NIH maintains substantive research classification systems: Computer Retrieval of Information on Scientific Projects (CRISP) assigns subdisciplinary classification terms to each grant and contract, Medical Subject Heading (MeSH) provides subject description and classification information for every publication indexed in the Medical Literature Analysis and Retrieval System (MEDLARS) and MEDLINE. MeSH identifies source of research support, subject, author, title, journal, data, and descriptors of the research for every indexed research article published since 1981. These databases are used for literature searches and for evaluation of NIH activities in a given research area.

NIH uses a special database, MEDLINE, for bibliometric analysis. This database contains records of all articles, notes, and reviews that have appeared since 1970 in a selected group of biomedical journals, along with the sources of financial support acknowledged by the authors of each article. There are over 300,000 papers in the database, as well as a record of nearly 2.5 million citations to those papers. Originally, the 240 journals of the database covered about 80 percent of the publications resulting from NIH-supported research. The size of the journal base was expanded in 1981 to include the entire MEDLARS system of the National Library of Medicine, nearly 1,000 journals, accounting for 95 percent of NIH-supported research. This extensive database has been

Helen Hofer Gee, "Resources for Research Policy Development at the National Institutes of Health," typescript, presented before the Health Policy Research Working Group, Harvard University, Mar. 20, 1985.

"Ibid.

the subject of the bulk of U.S. bibliometric studies, most of which seek to measure the long-term scientific payoffs from NIH-supported research.

Bibliometric Studies at NIH

Grace M. Carter of the Rand Graduate Institute conducted the first NIH commissioned bibliometric study in 1974.[25] Carter explored the use of citations as a measure of the research output of 747 research project grants and 51 program project grants awarded on a competitive basis in fiscal year 1967. The three output measures for research grants were the priority score received on renewal applications, the production of at least one frequently cited article, and the average citation rate for publications cited at least twice. Using statistical multivariate analyses, Carter tested a series of hypotheses about the three research output measures. Her examination of study section judgments of renewal applications revealed that on average grants proposed for renewal produced more useful research results than other grants. She also found that a grant that produced a highly cited publication was more likely to be renewed than one that did not. Thus, peer group evaluation and citation analysis produced comparable results.

In addition, Carter found a high correlation between priority scores on the first grant application and the number of subsequent publications and citations. She also found that research proposals perceived by study sections to have a high probability of being "exceptionally useful" received higher priority scores and more years of funding than those not so perceived. Carter concluded, rather cautiously, that the concept of "scientific merit" contains enough objective content that different groups of people meeting several years apart will agree that one set of grants is more scientifically meritorious than another set of grants.[26]

Francis Narin furthered the work of Carter by using bibliometric techniques to obtain quantitative indicators of research performance that were in general accord with the intuitive expectations of the research community .27 He was able to establish a degree of concordance between the structure of biomedical research literature and the structure of biomedical knowledge, which enabled him to use bibliometric databases and analyses to demonstrate a number of interesting points:

- Utilizing correlational techniques, he was able to establish correspondence between bibliometrically measured research productivity indicators and quantitative, nonbibliometric measures, including institutional funding and institutional ranking based on formal peer assessment.
- International biomedical publication rates are highly correlated with the GNP and national affluence (GNP per capita).[28]
- Changes in U.S. research funding can be associated with changes in the number and content of research publication 3 to 5 years later.
- Basic biomedical information is both published and cited by scientists supported by many bureaus, institutes, and divisions at NIH, forming a pool of fundamental research knowledge. In contrast, clinical information is produced and used by a narrower set of largely clinical researchers. Basic research is more highly cited than clinical research.
- Differences exist in the kinds of research publications produced by scientists in medical schools of different sizes and levels of national prestige. The number of publications

'Grace M. Carter, ***Peer Review, Citations, and Biomedical Research Policy: NIH Grants to Medical School Faculty,*** prepared for the Health Resources Administration and the Office of the Assistant Secretary for Planning and Evaluation of the Department of Health, Education, and Welfare, R-1583-HEW (Washington, DC, HEW, December 1974).

"Ibid., p. v.

'Francis Narin, ***Concordance Between Subject ve and Bibliometric Indicators of the Nature and Quality of Performed Biomedical Research,*** a Program Evaluation Report for the Office of Program, Planning and Evaluation, National Institutes 01 Health (Washington, DC: NIH, April 1983); Francis Narin, ***Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity,*** monograph prepared for the National Science Foundation, Accession #PB252339/AS (Springfield, VA: National Technical Information Service, March 1976).

[28]J. Davidson Frame and Francis Narin, "The International Distribution of Biomedical Publications," ***Federation Proceedings,*** vol. ***36, No. 6,*** May **1977, pp.** 1790-1795. Frame and **Natin** investigated the U.S. role in international biomedical publication based on counts of articles, notes, and reviews in 97s biomedical journals, Thev found that the United States authored 42 percent ot these papers which were tar more heavily cited than papers trom other countries. Only 4 percent of publications were found to originate trom underdeveloped regions.

produced per dollar of research funding is the same for the large and small institutions, indicating neither economies or diseconomies of scale. However, scientists from the larger medical schools publish their papers in more prestigious journals, and in a much wider set of subfields than smaller schools. Smaller schools can attain a critical mass of research activity only if they concentrate their research effort in a select area. In addition, faculty perceptions of the ranking of medical schools are very much in accord with bibliometric measures of the ranking of the same schools.

- The research supported and performed by the different institutes is appropriately concentrated in the clinical areas corresponding to their missions.
- Publications resulting from research conducted at NIH are more highly cited than publications in the same research areas supported through other sources.

Narin's work was the basis of the most widely accepted application of bibliometric techniques—the *Science Indicators* series of the National Science Foundation[29]—and provided the fuel for more extensive use of bibliometrics for evaluation and planning at NIH. Evaluations include the effectiveness of various research support mechanisms and training programs, the publication performance of the different institutes, the responsiveness of the research programs to their congressional mandate, and the comparative productivity of NIH-sponsored research and similar international programs.

Bibliometric analysts tested the ability of citation maps and co-citation clusters to detect transitions between basic research and clinical research in the biomedical sciences.[30] Citation maps of research in Lesch-Nyhan syndrome, Tay-Sachs disease, and the effects of the drug methotrexate displayed the anticipated transitions, though the extent of the transition varied from case to case. Co-citation cluster analysis identified the basic or clinical orientation of the different research areas but did not find the sought-for transition points.

Most recently, NIH has undertaken bibliometric studies to determine the effectiveness of different research support mechanisms. For example, a study of the research centers' programs using subsequent grant applications and publications as the criteria, found that a grant to a center is a more effective mechanism for supporting clinical research than it is for supporting basic science.[31] The Dental Institute used bibliometric analysis to determine that their centers' program has been effective in recruiting new scientists to research relevant to the institute's mission. Similarly, the National Cancer Institute and Francis Narin are conducting a study to identify the contributors to the most important research findings of the last 15 years and to determine where the research was conducted and the mechanisms of support. The research will test some assumptions of biomedical research grants policy, for example, that the individual investigator grant is superior to the contract and that extramural science is better than intramural.[32]

Bibliometrics have also been used for evaluating biomedical manpower training programs. Peter Coggeshall, a staff member of the National Academy of Sciences (NAS), used the NIH database, NSF grant data, and the NAS Survey of Doctoral Recipients to compare investigators who had received NIH predoctoral training support with two other groups—a group that had been trained in departments that had received training grants and a group that had received no NIH support in any form. The study concluded that individuals who received NIH predoctoral support produced superior subsequent career records in terms of publications and citations, were more likely to be working on NIH-sponsored activities, and were more successful in obtaining grants.[33]

Although Narin's early work showed that the distribution of papers supported by each institute

[29] *Science Indicators 1972; 1974; 1976; 1978; 1980; 1982; 1984*, **reports to the National** Science Board, National Science Foundation (Washington, DC: U.S. Government Printing Office, 1973, 1975, 1977, 1979, 1981, 1983, 1985)

[30] U. S. Department of Health and Human Services, *Applications of Bibliometric Methods* ***to the Analysis and Tracing*** *of Scientific* ***Discoveries,*** HHS-NIH Evaluation Report NTIS #PB80-210586 (Springfield, VA: National Technology Information Service, 1981).

[31] Gee, op cit, p. 9

[32] Lou Carrese, Associate Director for Program Planning and Analysis, National Cancer Institute, personal communication 1985

[33] Institute of Medicine, *The Career Achievements of NIH Predoctoral* ***Trainees and Fellows*** (Washington, DC National Academy Press, 1984).

in the basic and clinical medical disciplines follow very closely the institute's mission, several institutes continue to pursue the use of bibliometrics to validate accountability. For example, Narin has recently used bibliographic methods to evaluate trends in pulmonary and hypertension research. He found that National Heart, Lung, and Blood Institute actions since the passage of the National Heart, Blood Vessel, Lung, and Blood Act of 1972 have led to quantifiable progress in the research areas listed in the mandate .[34] Narin's work with the National Cancer Institute will be applied to the same purpose. In addition, the National Institute of Mental Health is conducting a 10-year analysis of the publication record of its grantees for purposes of accountability.

In almost all cases, bibliometric studies evaluated program performance and conformity to agency or institute mission. In some cases, they helped to identify areas for future research funding. The National Institute of Mental Health has begun to use bibliometrics and cluster groups to conduct a form of "portfolio analysis. " Looking at their program portfolios and the clustering of research publications by field, they are identifying leading edges of research that might require more support from their institute. 35 Narin has shown empirically that bibliometric data may qualify as an important adjunct measure to more subjective measures applied through peer review.

The Utility of Bibliometrics for Research Decisionmaking

Bibliometric techniques provide rough indicators of the quantity, impact, and significance of the output of a group of scientists' research. They are not generally considered valid for measuring the productivity of individual scientists due to differences in publishing styles and journal requirements, and the questionable validity of small statistical samples. However, Cole and others have shown that publication counts correlate positively with other measures of individual scientists' research quality such as peer review, Nobel Prizes, and prestige of academic appointment .36

Publication counts give a rough measure of the quantity of work produced by a research team or facility. Citation counts are an indicator of the influence that work has had on the larger scientific community. And the number of citations per article or the number of highly cited articles provide a rough measure of the significance of the work, since important papers tend to be cited most often. These indicators can help a funding agency compare the quantity, quality, and visibility of research done by various individuals or institutions. They can help identify the strong research groups and the relative cost-effectiveness of research sponsored at different centers.

However, they have two important limitations with respect to research decisionmaking. First, they are entirely retrospective. They have no inherent future predictive capability, unless one believes that past performance is an indicator of likely future achievement—not an unreasonable assumption. Therefore, they are more applicable to research program evaluation than to research planning. Second, they are not applicable to *strategic* decisions about resource allocation between fields. Most bibliometricians contend that the techniques can only be validly applied within individual disciplines. Publication and citation practices vary too widely between fields to allow for interdisciplinary comparisons. This, unfortunately, makes bibliometric techniques of *limited value for the most important decisions facing agency heads and congressional decisionmakers—allocating resources among fields.*

It should be noted that some analysts dissent from this view. Derek de Solla Price, in an unpublished article for the National Research Council, argues that the relative strength of different

'Public Health Service, ***Bibliographic Methods for the Evaluation of Trends* in *Pulmonary and Hypertension Research,*** NTIS #PB82-159724, 1982.

"Lawrence J. Rhoades, Science Policy Planning and Evaluation Branch, Office of Policy Analysis and Coordination, National institute of Mental Health, personal communication, 1985.

'G. A. Cole, ***The Evaluation of Basic Research in industrial Laboratories*** (Cambridge, MA: Abt Associates 1985) p 43 fields in the United States can be assessed by comparing the ratios of the numbers of citations of U.S. articles in foreign journals to the numbers of citations of foreign articles in U.S. journals by field, normalized to account for differences in national research "output" by field. Price's scheme is quite complicated and involves measures of quality, quantity, and "internationality, " but it is a first attempt to compare fields using dimensionless indicators that have been normalized to remove the effects of different publishing and citing practices between fields. 37

Co-citation analysis enables one to monitor how specialties or subfields evolve over time. Cogitation analyses display the relationships among highly cited papers by *showing how many times such papers are cited together in single articles.* Based on co-citations, two-dimensional diagrams or maps of specialties can be created which illustrate the clustering of the most important works in that specialty, based on the number of citations By examining changes in the clusters one can track the evolution of the specialty over time. For ex - ample, figures 2A, B, C, and D illustrate the evolution of the collagen specialty cluster in biochemistry between 1970 and 1973. As can be seen by comparing figures 2D and A, the cluster map has become much larger by 1973, and most significantly, an entirely new set of research papers has replaced the cluster of most important works identified in 1970. This change coincided with the discovery of a new substance, pre-collagen, in 1971, which totally reoriented the research front in the specialty. Thus co-citation maps can, in principle, help one to identify important changes in research specialties over time.

‘-Derek de Solla Price, "Science Indicators ot Quantity and Quality for Fine Tuning of United States Investment ın Research ın Major Fields of Science and Technology, " typescrıpt draft, paper prepared at the request of the Commission on Human Resources, National Research Council, April 1980, typescript draft.

Even without maps, co-citation cluster analysis can help one identify the level of research activity in different specialties. Table 10 takes specialty

Table 10.—Changes in Sample of Continuing Clusters, 1970-73

Specialty	Direction of change	1970-71 (%)	1971-72 (%)	1972-73 (Ye)
Nuclear levels	c	58	45	25
	d	21	55	17
	n	21	0	58
Adenosine trlphosphatase	c	67	25	67
	d	0	50	22
	n	33	25	11
Australia antigen	c	55	54	57
	d	4	26	30
	n	41	20	13
Proton-proton elastic scattering	c	50	7	44
	d	50	21	25
	n	0	72	31
Ultrastructure of secretory cells	c	50	43	60
	d	12	57	0
	n	38	0	40
Nuclear magnetic resonance	c	37	55	23
	d	13	9	54
	n	50	36	23
Polysaccharides	c	46	44	36
	d	46	34	7
	n	8	22	57
Crystallizatlon of polymers	c	100	100	100
	d	0	0	0
	n	0	0	0
Affinity chromatography	c	60	67	72
	d	20	0	14
	n	20	33	14
Leukocytes: chonic granulomatous disease	c	40	63	33
	d	13	5	53
	n	47	32	14
Collagen	c	80	40	27
	d	20	0	40
	n	0	60	33
Erythrocyte membranes	c	9	15	58
	d	64	5	42
	n	27	60	3
Delayed hypersensitivity	c	77	46	50
	d	15	27	29
	n	8	27	21

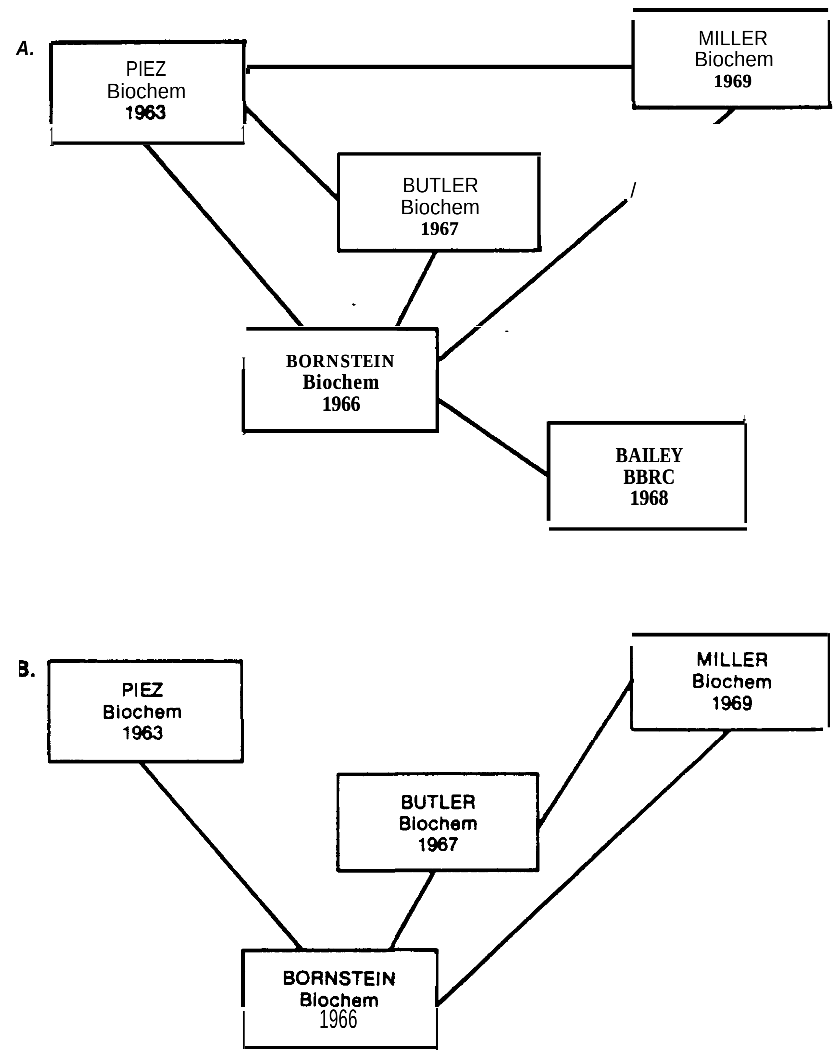
c - continuing, d = dropping, n - new documents

SOURCE Yehuda Elkana, et al., (eds) Toward a *Metric of* Science *The Advent of Science Indicators* (New York: John Wiley & Sons. 1978), pp 199-201 clusters in biochemistry and shows, for each, the percentage of key papers that are identical to the previous year's, the percentage that have dropped out from the previous year, and the percentage that are new. As can be seen, the specialties vary appreciably, from crystallization of polymers, for which the same papers defined the subfield cluster over all 3 years, to erythrocyte membranes, in which 64 percent of the important papers dropped out in 1970, and 80 percent of the papers were totally new in 1971. If research decisionmakers are eager to fund specialties where new ideas are emerging rapidly, data on the evolution of cluster specialties over time could help to identify fields of rapid change.[38]

It should be stressed, however, that most bibliometncians view publication, citation and co-citation analyses as complements to, not substitutes for, informed peer evaluation. All three analyses are, of course, ultimately indirect measures of the scientific community's peer evaluation of researcher's productivity.

"Eugene Garfield, et al., "Citation Data as Science Indicators, *Toward* **a** ***Metric of Science: The Advent of Science indicators,*** Yehuda Elkana, et al., (eds.) (New York: John Wiley & Sons. 1978), pp. 196-201.

Figure 2.–Davelopment of a Speciality Cluster, 1970-73

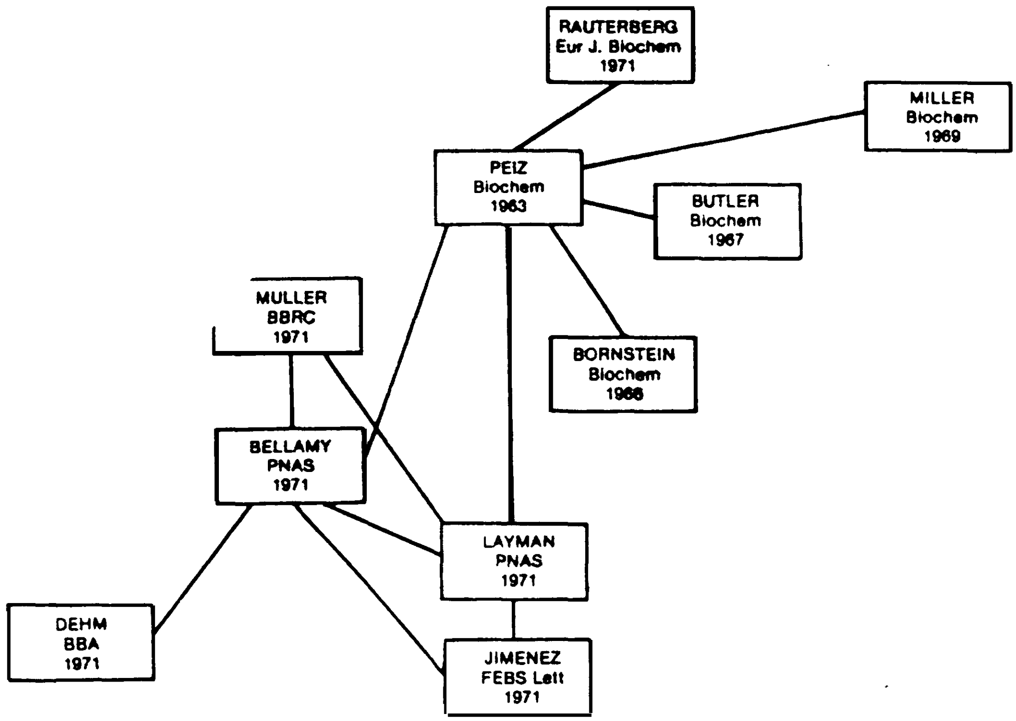


The figure shows the evolution of the collagen cluster over the 4-year period 1970-73. Boxes contain the names of first authors of the highly cited papers and years of publication. Lines connect papers co-cited at least 11 times in the corresponding source year.

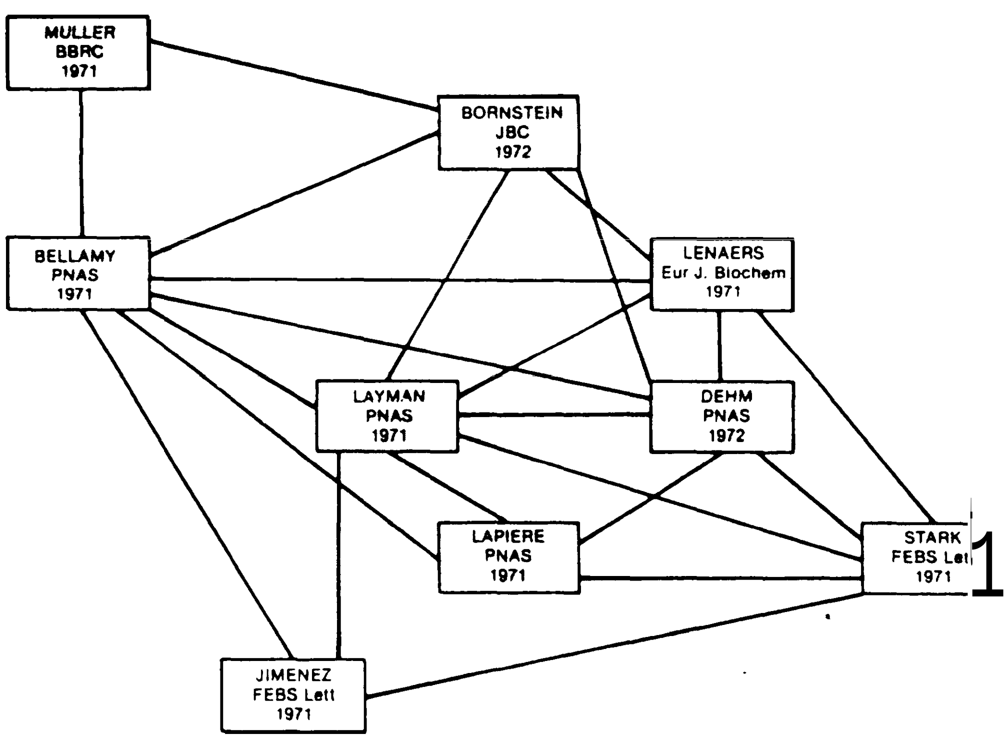
NOTE A IS collagen, 1970, and B IS collagen, 1971

Figure 2.—Deveiopment of a Specialty Cluster, 1970-73 (continued)

C.



D.



NOTE: C. is collagen, 1972; and D. is collagen, 1973.

SOURCE: Eugene Garfield, et. al., "Citation Data as Science Indicators," *Toward a Metric of Science: The Advent of Science Indicators*, Yehuda Elkana, et al., (eds.) (New York: John Wiley & Sons, 1978), pp. 199-201.

SCIENCE INDICATORS

One method of assessing the health of the research enterprise is to directly question the scientists and administrators involved in it. This is done in a variety of ways in this country. Individual scientists are asked to testify at congressional hearings. Federal agencies create scientific advisory panels to help guide research.

The National Research Council and its constituent bodies—the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine—carry out numerous reviews of research programs and research fields for the executive and legislative branches. The most comprehensive of these are the Research *Briefings* and Five *Year Outlooks* prepared by the Committee on Science, Engineering, and Public Policy. Since all NRC reports are prepared by committees of scientists, they represent, to some degree, the informed, consensus-based peer judgments of the scientific community on the state of the research enterprise. As a check on the validity of such reports, the government could support detailed surveys of the scientific community.

However, scientists' judgments on the state of their own field of research can rarely be totally disinterested and often reflect the researcher's characteristic desire to investigate more problems in greater depth than available funding will allow. These inherent biases can be compensated, to some degree, by the use of a variety of science indicators that are measured by NSF, NIH, and NRC and published on a regular basis. These indicators include the amount of funds devoted by the Nation to research and development by source, sector, nature of the work, performer, and scientific fields; statistics on the distribution of scientific and engineering personnel, graduate students, and degree recipients by field, sector, and institution; and the support for graduate education and training. The NSF Science *Indicators* tables include funding levels by agency and even program; the specific institution receiving the funds, employing the scientists, and training the graduate students; and funding for specific Standard Industrial Classification codes within industry. Most lacking in the policy community is a consensus on which indicators are most relevant and how the different indicators might be used in combination to measure the health of the research enterprise. A workshop or report on the use of Science *Indicators* to measure the health of the research enterprise might be a useful first step in that direction.

One must remember, however, that all measures or "indicators" of research are inevitably flawed. Any number describing research is an abstract symbol that depicts, imperfectly, only one aspect of it. Choosing one measure over another implies that the measurement use has made some assumption about what is important. The chosen measure has meaning only through interpretation.

Even if an acceptable measure of an aspect of research can be devised, interpretation remains problematic:

> . . . the inputs [to science]—of dollars, of working scientist, males, females, and Hispanics, graduate students, post-docs, and professors—are well known and further broken down into industry, government, education, or lost to view. We also have counts of outputs—of papers, citations of papers, and Nobel Prizes arranged according to national origin. But how do we know what the numbers "ought" to be? . . . such indicators help very little in determining the health of science in any absolute sense or, more practically, in relation to what it might be if organized and financed at some theoretical optimal level .[39]

These difficulties illustrate the limitations of science indicators for research evaluation.

[39] R.S. Morison, "Needs, Leads, and Indicators, " Science, *Technology, and Human Values,* vol. **7, winter 1982,** pp. 6-7.

Chapter 4

Research Decisionmaking in Industry: The Limits to Quantitative Methods

Chapter 4

Research Decisionmaking in Industry: The Limits to Quantitative Methods

This chapter reviews the use of quantitative analysis in industry research and development (R&D) decisionmaking at the level of the individual firm. OTA reviewed articles, surveys, and reports, and interviewed several research managers. Very little systematic information is available about industry's use of quantitative models in research resource allocation. A few surveys cover limited numbers of firms and are not necessarily representative. Much of the material is in the form of anecdotal accounts of an individual firm's uses of certain methods, providing no information on whether other firms have adopted similar approaches. Relatively more information is available about noneconomic quantitative methods for project selection than about economic models. The literature, as well as OTA interviews, demonstrates the limited practical utility of quantitative techniques for research decisionmaking in industry, and the reliance on subjective judgment and good communication between R&D, management and marketing staffs in the decisionmaking process.

In the private sector, R&D is an investment that must compete for corporate support with other investment opportunities such as plant expansion or new product marketing. Program and laboratory directors must defend the value of their research to top management and decide what mix of projects is best for the firm. Project managers must determine whether their projects are proceeding as planned and whether expected payoffs will justify costs. This chapter looks in turn at the use of quantitative methods in review and evaluation of ongoing research, new research project selection, and research resource allocation as part of strategic planning.

REVIEW AND EVALUATION OF ONGOING RESEARCH ACTIVITIES

As part of their normal operations and management, firms periodically review research programs, projects, and staff to assess progress and determine the contribution that individual researchers and research groups are making to the firm's goals. These reviews can justify research expenditures to management, assist in budget and. program planning, or evaluate personnel performance.

Few firms use quantitative methods to review ongoing research. In 1982, Schainblatt surveyed 34 R&D-intensive firms about methods of measuring research productivity.[1] The survey focused on the groups, programs, or other organizational units, rather than on individual scientists or engineers. Managers in only four firms reported using performance or output measures as part of their program reviews. Only 20 percent of the firms routinely collected any kind of productivity data, and 20 of the 34 firms reported using no productivity-related measures at all.

Several respondents said they had tried for years to measure R&D productivity but had not been successful. Noted one research manager, "We . . . came to the conclusion that there is no good way to do it on a week-to-week or month-to-month basis. " Managers doubted that R&D productivity measures were meaningful. According to one manager, "Attempts to quantify benefits of R&D have led to monstrosities that caused more harm than good. "[2]

A. Schainblatt, "How Companies Measure the Productivity of Engineers and Scientists, " *Research Management,* May 1982, pp. 10-18.

'Gerald A. Cole, *The Evaluation of Basic Research in Industrial Laboratories* (Cambridge, MA: Abt Associates, 1985), p. 59.

Macroeconomic Models

The retum-on-investment (ROI) model is one of a family of economic analysis tools, along with discounted cash flow analysis and present value analysis, which business managers use as aids to investment decisions. These methods are commonly applied to decisions where uncertainty is low. As we have seen in chapter 2, research investment decisions entail considerable uncertainty. Moreover, the ROI methods carry a bias against the long-term, high-risk, and high-uncertainty projects like basic research, and work in favor of high-vield, short-term investments.[3] The bias increases in periods of high inflation like the 1970s.

Despite the fact that economists have typically viewed R&D activity as an investment, a number of scholars recently have criticized the misuse by private sector managers of quantitative financial techniques for evaluating investments in R&D and technology more generally. Hayes and Abernathy argue that the application to investment decisionmaking of principles of portfolio management has led U.S. firms to underinvest in new technologies:

> Originally applied to help balance the overall risk and return of stock and bond portfolios, these principles have been applied increasingly to the creation and management of corporate portfolios —that is, a cluster of companies and product lines assembled through various modes of diversification under a single corporate umbrella. When applied by a remote group of experts primarily concerned with finance and control and lacking hands-on experience, the analytic formulas of portfolio theory push managers even further toward an extreme of caution in allocating resources.[4]

Similarly, Hayes and Garvin have argued that present-value analysis of R&D investment decisions has led to a systematic bias against such investments, due in part to the use of discount or "hurdle" rates for such decisions that are too highs Rather than relying solely on quantitative techniques for the evaluation of R&D investments, these authors argue that managers must develop an understanding of the underlying technologies, and apply informed judgment in making such decisions.

ROI methods, as applied to R&D projects, estimate the economic value of research and compare it with the cost to the organization. For example, a firm could estimate the sales or revenues generated or expected from a new product resulting from research efforts. Alternatively, the firm could estimate the share of total profits or savings attributable to research. The financial results are usually discounted to reflect the time value of money.

One major problem with such methods is that it is difficult to apportion profits generated by a product developed in the past among research, development, and marketing activities, all of which contribute to the process. Even if one could attach accurate figures to the present payoff from past research, such calculations offer little guidance for current decisions, which are occurring under different technological, management, economic, and organizational conditions.[b] In addition, companies often acquire R&D through purchase, corporate acquisition, or merger; or they may sell R&D themselves.

ROI techniques seem most applicable to justifying the value of ongoing research to top management by placing a value on past research. Such

---

[3] J.E. Hodder and H.E. Riggs, "Pitfalls in Evaluating Risky Projects," *Harvard Business Review*, January-February 1985, pp. 128-135; and G.F. Mechlin and D. Berg, "Evaluating Research—ROI 1s Not Enough," *Harvard Business Review*, September-October 1980, pp. 93-99.

[4] Robert H. Hayes and William J. Abernathy, "Managing Our Way to Economic Decline," *Harvard Business Review*, vol. .s8, No. 4, 1980, pp. 67-77, reprinted in *Survival Strategies for American industry*, Alan M. Kantrow (cd.) (New York: John Wiley & Sons, 1983), pp. 15-3S (quotation is from pp. 22-23, 1983 reprint).

[5] As these techniques have gained ever wider use in Investment decisionmaking, the growth of capital Investment and R&D spending m this country has declined We summit that the discounting approach has contributed to a decreased willingness to invest for the following reasons: (1) it is often based on misperceptions of the past and present economic environment; and (2) it is biased against investment because of critical errors in the way this theory is applied Bluntly stated, the willingness of managers to view the future through the reversed telescope of discounted cash flow is shortchanging the future of their companies.

Robert H. Hayes and David A. Garvin, "Managing As If Tomorrow Mattered," *Harvard Business Review*, VOI. 00, No. 3, pp. 70-79, reprinted in *Survival Strategies for American Industry*, Alan M. Kantrow (cd.) (New York: John Wiley & Sons, 1983), pp. 30-51, (quotation is from p. 37, 1983 reprint).

[6] B. Twiss, *Managing Technological Innovation* (New York Longman, 1980), pp. 121-122; and D W. Collier, "Measuring the Performance of R&D Departments," *Research Management*, VOI, 20, No. 2, 1977, pp. 30-34.

techniques do not appear very helpful for determining changes in a research program's budget, or deciding the fate of particular project. The ROI methods and other cost/benefit methods, as demonstrated below, are more applicable to decisions involving setting priorities among a set of applied research projects.

Business Opportunity Techniques

Collier describes a "business opportunity" technique that avoids some of the pitfalls of ROI appreaches. ' It is based on the notion that the primary objective of industrial research is to identify and define business opportunities that can be exploited commercially. Management evaluates the performance of research staff by comparing its technical accomplishments against a set of previously established objectives—for example, to develop a new control system with certain performance characteristics. Next, when the project is completed and is ready to be transferred to the production and marketing departments, total expected sales revenues are estimated and discounted to the present. The result, when divided by the project's cost, is a return on research figure which, if summed across all completed projects, yields an estimate of the research department's value to the company for that year.

The business opportunity method depends on the accuracy of future sales estimates and of product development time, which are subject to substantial error. Thus, uncertainties in the ROI method's allocation of credit for profits retrospectively among the various factors are replaced by the business opportunity method's uncertainties about future payoffs. *Collier evidently assumes that there is a higher level of certainty about future profits than about the contribution to profits of past activities.* Nevertheless, Collier states explicitly that neither the ROI nor the business opportunity method should be used to evaluate basic research, ' probably because particular basic research projects and programs are difficult to tie uniquely to economic impacts, which may well be separated from the research by many years and institutional boundaries.

Little is known about the use of other formal evaluation techniques such as bibliometrics, patent counts, and colleague surveys for peer review. DuPont, Bell Laboratories, and other large industrial research establishments carry out intensive, annual reviews of their scientists' work. These reviews stress the scientists' contribution to science and technology, particularly in areas of strategic interest to the company. The reviews are performed by peers in the fields of research concerned.[9] No evidence suggests that this practice is widespread. One complicating factor is that much industrial R&D is shared work, so attributing some portion of its output to one individual is difficult.

'Ibid.
'Cole, op cit., p. **38.**

'Ibid., p. 59

R&D) PROJECT SELECTION

Relatively more information exists about industry's use of formal, quantitative techniques to select R&D projects and shape research portfolios. An extensive literature on models and methods exists and it includes some surveys on their use. These algorithms or heuristic devices help managers assign values to projects, groups of projects, or other investments. The approaches fall into four categories:

1. scoring models,
2. economic models,
3. constrained optimization or portfolio models, and
4. risk analysis or decision analysis models.

Scoring Models. When scoring models are used, each project is rated against a series of relevant decision criteria. Scores for each project are combined through addition or multiplication to develop a single project score.

In a typical application, all candidate projects are scored and ranked from highest to lowest.

Since costs are associated with each project, the allocation decision involves simply going down the list of projects until the available funds have been exhausted. This procedure does not consider the effect of variations in project budgets, marginal returns from varying project funding levels, interactions among projects, or changes in annual budget levels over the life of the projects.

Scoring models have the least demanding input data requirements of the four categories of models. They are designed to incorporate noneconomic criteria, and can operate on input data in the form of subjective estimates from knowledgeable people. The assumptions underlying scoring models are relatively undemanding: only that explicit evaluation criteria and a way to quantify each evaluation be developed. While expert judges may be used as the source of qualitative input data, eventually these data must be expressed quantitatively to be used in a scoring model. The choice of algorithm to convert qualitative data to quantitative scores is arbitrary and subjective.

Economic Models. With economic models, projects are rated against a series of economic criteria such as expected rate of return. A single figure of merit is produced, typically reflecting the ratio of the present value of earnings from the project, including the probability of project success, to discounted money flow or investment. These are essentially capital budgeting models. Economic models accept only quantitative data based on estimates of the financial performance of the project over a specified planning horizon. These estimates are often generated by program or project managers or by panels of experts. They possess no greater intrinsic validity than data developed for scoring models. This is particularly true when uncertainties about the technical and market performances of the technology are high, a situation frequently encountered in early stages of applied research. Any estimate of future project benefits requires subjective input from some well-informed respondent or group of respondents. [10]

Like scoring models, economic models produce a single figure of merit that is independent of the figures of merit for competing projects. The simplest application is to rank projects and fund those scoring highest until the available funds are exhausted. The narrow focus of economic models may limit their usefulness for government research, where cost and economic gain are only part of a much larger set of criteria.

Constrained Optimization or Portfolio Models. These models measure a program's potential to meet a goal, usually a series of economic objectives, subject to specified resource constraints. Unlike scoring and economic models, the focus is on a mix of projects rather than a simple ranking of individual projects. In portfolio analysis, mathematical programming techniques are used to evaluate the allocation of resources among candidate projects. The method requires an understanding of the relationship among resource inputs, technological performance, and marketplace response. The decisionmakers must agree that funding just the right mix of resources among projects is the key to effective R&D management.

While demanding better data quality and understanding of underlying technical and economic processes, portfolio models can handle multiple constraints and different budget levels over different years in the planning horizon. But the use of these models implies a level of management control and flexibility to reallocate resources that may not exist in many situations, particularly in government. [11]

Risk Analysis or Decision Analysis. These models produce an expression for the expected utility of each set of alternative budget allocations among a set of research projects. Models using decision analysis have the most complex data requirements, since inputs must be in the form of probability distributions. As in the case of portfolio analysis, considerable understanding of underlying processes must exist if the benefits of decision analysis are to be realized. Decision analysis incorporates expert judgments as well as "objective" data.

N'. R. Baker, "R&D Project Selection Models: An Assessment, " ***IEEE Transactions on Engineering Management,*** EM-21, November 1974.

'K. G. Feller, *A Review of Methods for Evaluating R&D* (Livermore, CA: Lawrence Livermore Laboratory, 1980), p. 19

Use of R&D Project Selection Models in Industry

A 1964 survey of the use of quantitative R&D project selection methods concluded that while numerous models and techniques had been proposed, available data showed "little thorough testing and only scattered use of the proposed methods."[12] A second survey published 2 years later led to a similar conclusion:

> The practice of project selection in industry and government is dominated by . . . methods depending heavily upon individual or group judgment and using very little quantitative analysis. The use of cost and return estimates is common, but very few organizations employ any formal mathematical model for combining these estimates and generating optimal project portfolios.[13]

Rubenstein pointed out in 1966 that the use of quantitative methods by R&D organizations had not increased appreciably since 1950, and that the reasons for this had more to do with the nature of the available data and the R&D decision process than with the sophistication of available models.

In a 1968 study of the R&D project selection practices of 36 firms, Dean found that formal, quantitative models were not widely used. Simple scoring models employing only a few criteria such as probability of technical success, estimated time to completion, cost, and size of net market gain were the only mathematical models that had been tried.[14] Meadows' 1968 report on practices in the R&D labs of five major companies found that the margin of error in estimates undermined the usefulness of the methodologies.[15] He provides a telling example:

> If estimates with only a ten percent error were inserted in the formula, they could conceivably lead management to calculate a higher profit ratio for a project actually expected to lose money for the firm than for one expected to return 230 percent on the money invested in its development. This sensitivity of the model's output to error is important in view of the fact that no laboratory yet studied has had estimates (of error) averaging as little as ten percent.[16]

Mansfield, who has conducted numerous studies of R&D and innovation in industry, noted in a more recent article that:

> . . . most companies . . . have found it Worthwhile to make economic evaluations of project proposals and continuing projects, often adapting such capital budgeting techniques as rate of return or discounted cash flow to the task at hand.[17]

But he goes on to say that the nature of the techniques used will vary, depending on the stage of research. Early on, when costs are low and uncertainty high, project screening will be quick and informal. Later, the larger labs make some use of quantitative methods:

> In some labs, they [quantitative methods] are taken quite seriously indeed; in others they are little more than window dressing for professional hunches and intra-company politics . . . The more sophisticated types of models have not been extensively used.[18]

Although it is important to distinguish among the various research activities when discussing the use of such models, surveys rarely do so. One notable exception is a 1971 study of project selection practices by a task force of the Industrial Research Institute. The task force studied 27 companies and classified their R&D programs into three types: exploratory, high risk business development, and support of existing business,[19] Among the firms studied, those engaging in exploratory R&D generally used simple, unsophisti-

---

[12] N. R. Baker and W.H. Pound, "R&D Project Selection: Where We Stand," *IEEE Transactions on Engineering Management,* EM-21, November 1974, p. 130.

[13] A.H. Rubenstein, "Economic Evaluation of Research and Development: A Brief Survey of Theory and Practice," *The Journal of Industrial Engineering,* **vol. 17, November 1966, p. 616.**

[14] B. V. **Dean, "Evacuating, Selecting and Controlling R&D Projects,"** ***Research Study 89*** **(*New*** York: American Management Association, 1968).

[15] D. L. Meadows, "Estimate Accuracy and Project Selection Models in Industrial Research," *industrial Management Review,* spring 1968, pp. 105-119

[16] 1 bid., p. 116.

[17] Edwin Mansfield, "How Economists See R&D" *Research* ***Management,*** ***vol. 25, 1982,*** p. 25

[18] Ibid.

[19] R, E. Gee, "A Survey of Current Project Selection Practices *Research Management,* ***vol.*** 14, September 1971 pp 38-45

cated selection procedures. Decisions on funding were based on a page or two of qualitative information or a simple rating scheme. Decisions on high risk business development projects were occasionally supported by more sophisticated, quantitative techniques such as standard economic projections. There was very limited use of quantitative methods for dealing with uncertainty. For decisions about projects in support of existing business, on which quantitative data with very low uncertainty could be brought to bear, standard economic projections were widely used. A 1985 survey indicates little has changed: industry managers rely on qualitative evaluation of basic research programs and proposals.

There is substantial agreement that the less complex scoring models with less demanding input requirements are more appropriate for earlier stages in the R&D process than the more analytically sophisticated economic and linear programming models. One literature review concluded that economic models are too quantitative for evaluating even applied research efforts; they can help only in identifying the information needed to make a qualitative estimate of the economic merit of applied research projects.[20] Twiss, in the second edition of his respected text on the management of technological innovation, finds little value in sophisticated analytic models for R&D project selection:

> While the formulae may give satisfactory symbolic representation, it is doubtful whether they provide a mechanism of much operational value. The judgments involved are so complex there is a great danger of the formulae being used to apply a veneer of pseudo-quantification to support decisions which have already been taken on different considerations . . . If the data is poor they are little better than descriptive representations of the problem. When applied to estimates of the order of inaccuracy discussed earlier, they can do a positive disservice by concealing in a simple index the magnitude of uncertainties. However, there are some types of R&D work where it is possible to assess both the benefits and costs to a high degree of accuracy. These are usually development projects.[21]

The basis for judging the appropriateness of different types of project selection models for different stages of R&D should not rest on distinctions between quantitative and qualitative data, since it is always possible to assign a number to qualitative data using some arbitrary algorithm. Rather, the association should be based on the level of certainty involved in the estimates of the probability that technical and market performance goals will be achieved at a certain cost within a specified time. In basic research and in the early stages of applied research, these factors can be predicted with little certainty. (See box B.)

Reasons for Levels and Patterns of Observed Use

Industry managers recognize that attempts to link basic research activities directly and quantitatively to any kind of "payoff"—new products, profits, corporate image internal consulting, scientific knowledge, or personnel recruitment—are flawed and of limited value. The uncertainties are too great, the causal paths too diffuse, the benefits too difficult to measure, and the timeframe too extended. Basic research usually represents a small fraction (5 to 10 percent) of R&D budgets, and is not subject to the same financial scrutiny as applied research and development activities.

The realities of the research process account for the limited use of quantitative project selection models. The increasing sophistication of the models has not improved their acceptance. In fact, the increased sophistication may create as many limitations as it removes.[22] Fundamental inadequacies in the data required greatly limit their value. Studies of company estimates of project cost and time requirements show that they are usually highly inaccurate. Mansfield, et al.,[23] found that in one drug firm, the average ratio of actual to estimated development costs exceeded

[20] W.M. Burnett and D.J. Monetta, ***Applied Research Project Selection in Mission-Oriented Agencies**: An Approach* (Washington, DC: U.S. Department of Energy, Assistant Secretary for Energy Technology, Division of Power Systems, 1978).

[21] Twiss, op. cit., p. 135.

[22] E. P. Winkofsky, et al., "R&D Budgeting and Project Selection: A Review of Practices and Models," *TIMS Studies in the Management Sciences*, vol. 15, 1980, p. 192

[23] Mansfield, op. cit

Box B.—Differences Between Implicit Assumptions of Project Selection Models and Typical Decision Environments

Implicit Assumptions	Typical Decision Environment
1. A single decisionmaker in a well-behaved environment	1. Many decisionmakers and many decision influencers in a dynamic organization.
2. Perfect information about candidate projects and their characteristics; outputs, values, and risks of candidates known and quantifiable.	2. Imperfect information about candidate projects and their characteristics; project outputs and values are difficult to specify; uncertainty accompanies all estimates.
3. Well-known, invariant goals.	3. Ever-changing, fuzzy goals.
4. Decisionmaking information is concentrated in the hands of the decisionmaker who has all the information needed to make a decision.	4. Decisionmaking information is highly splintered and scattered piecemeal throughout the organization, with no one part of the organization having all the information needed for decisionmaking.
5. The decisionmaker is able to articulate all consequences.	5. The decisionmaker is often unable or unwilling ,to state outcomes and consequences.
6. Candidate projects are viewed as independent entities to be evaluated on their own merits.	6. Candidate projects are often technically and economically interdependent.
7. A single objective, usually expected value maximization or profit maximization, is assumed and the constraints are primarily budgetary in nature.	7. There are sometimes conflicting multiple objectives and multiple constraints, and these are often noneconomic in nature.
8. The best portfolio of projects is determined on economic grounds.	8. Satisfactory portfolios may possess many noneconomic characteristics.
9. The budget is "optimized" in a single decision.	9. An iterative recycling budget determination process is used.
10. A single, economically "best" overall decision is sought.	10. What seems to be the "best" decision for the total organization may not be seen as best by each department or party, so that many conflicts may arise.

SOURCE: Adapted from William E. Souder, "A System for Using R&D Project Evaluation Models," *Research Managem ent*, September 1978, pp. 29-37.

2 to 1 development time required exceeded estiates by a factor of almost 3. A more recent study of major innovations developed during a 5-year period by a large U.S. company indicated that initial estimates of an R&D project's expected profitability were no more reliable than the drug firm's cost and time estimates. "The chances were about 50-50 that the estimated discounted profit from a new product or process would be more than double or less than half the actual discounted profit ."[24] The inaccuracies of such measures point to the potential unreliability of quantitative evaluation of basic research activities.

The reasons for lack of reliance on models for research decisionmaking include:

- inadequate treatment of multiple, often interrelated criteria;
- inadequate treatment of project interrelationships;
- lack of explicit recognition and incorporation of the experience and knowledge 01 the researchers and managers;

"Mansfield, op. cit., p. 26

- inability to recognize and treat nonmonetary aspects of research programs that are difficult to understand and use; and
- inadequate treatment of program and staff evolution.[25]

[25]Winkofsky, et al., op. cit., pp. 191-192.

Mansfield adds that many models fail to recognize that R&D is a process of uncertainty reduction—in effect, buying information.[26] Thus, technical failures are successes in that they provide valuable information.

[26]Mansfield, op. cit., p. 25.

STRATEGIC PLANNING AND RESOURCE ALLOCATION

Throughout the 1960s, company managers generously funded "open-ended" research—research that was not necessarily directed toward increasing corporate profits. However, corporate investment in R&D decreased significantly in the 1970s. Dr. Alan Frohman, a management consultant and faculty member at Boston University's business school, believes management's attitude towards investing in R&D:

> . . . has seesawed from unquestioned support and optimism in the 1960s to withdrawn support and discouragement . . . in the 1970s.
>
> " In the 1960s, the attitudes were evidenced by the building of large, well staffed laboratories, often remote from the businesses. During the 1970s, the major, painful cutback in both expenditures and staffing documented management's discouragement with the contribution of technology to the bottom line.[27]

Giorgio Petroni, a professor who has studied the history of management, contends that this "discouragement" was caused by the lack of attention given to strategic planning.

> During the 1970s, management, even in "technology intensive" enterprises, showed little understanding of the need to develop technological expertise within their organizations. Top management often did not understand the full importance of technology as an element of competitive strategy.[28]

Petroni and Frohman are not the only scholars to cite the failure of management to plan for technological innovation as a major factor in the reduction of profits in American industries. In fact, in 1980, when Frohman's study was published, several major literature reviews emphasized what Alan Kantrow has called "the strategy-technology connection."[29]

After reviewing the research literature from the 1970s, Kantrow concluded that there is no rational justification for separating technology from strategy:

> Technological decisions are of fundamental importance to business and, therefore, must be made in the fullest context of each company's strategic thinking. This is plain common sense. It is also the overwhelming message of this past decade's research.[30]

Based on the little that was known, Kantrow tentatively identified the key elements of corporate technology strategy:

> . . . good communications, purposeful allocation of resources, top-level support within the organization, and careful matching of technology with the market.[31]

While scholars and managers alike knew little about technological planning in 1980, they knew a good deal about the theories and practices associated with strategic management. In the same issue of the *Harvard Business Review* that featured Kantrow's article, a review by Frederic W. Gluck, et al., stated: "for the better part of this decade, strategy has been a business buzzword."[32] The in-

[27]Alan L. Frohman, "Managing the Company's Technological Assets," ***Research Management,*** September 1980, pp. 20-24.

[28]Giorgio Petroni, "Who Should Plan Technological Innovation?" ***Long Range Planning, vol. 18, No. 5, 1985,*** pp. 108-115.

[29]Alan M. Kantrow, "Keeping Informed: The Strategy-Technology Connection," ***Harvard Business Review,*** July-August 1980, pp. 6-21.

[30]Ibid., p. 6.

[31]Ibid, p. 11.

[32]Frederic W. Gluck, et al., "Strategic Management tor Competitive Advantage," ***Harvard Business Review,*** July-August 1980, p. 154.

crease in the use, and misuse, of this term is important because it signifies a shift in managers' planning approach from technicalities "to substantive issues affecting the long-term well-being of their enterprise."[33] Despite this change, most managers did not apply these long-term strategies to their R&D divisions in the 1970s.

However, it is important to note that during this period there was a significant countertrend.[34] The verbose title of a 1973 *Chemical Week,* tells the story: "Research Gets the Word: If It Doesn't Fit, Forget It—It's The New Way of Life: R&D Must Mesh Closely With Corporate Goals."[35] In 1976, Union Carbide's R&D director told a reporter: "R&D is too important to be left to the R& D'ers; R&D is the future analog of today's capital expenditures."[36] The vice president of Celanese agreed: "At the high cost of R&D, we can no longer afford to plan and manage it in a random manner. It has to be very closely tied to strategic business planning."[37]

Eastman Kodak, a pioneer in industrial research, was one of the first companies to make a concerted effort to incorporate R&D issues into its strategic planning process in the 1970s. This represented a major change in the company's R&D policy from the 1960s when research director C.E. K. Mees articulated Kodak's hands-off policy:

> The best person to decide what research work shall be done is the man who is doing the research, and the next best person is the head of the department, who knows all about the subject and the work; after that you leave the field of the best people and start on increasingly worse groups, the first of these being the research director, who is probably wrong more than half of the time; then a committee, which is wrong most of the time; and finally, a committee of vice-presidents, which is wrong all the time.[38]

In 1978, Kodak formalized the relationship between corporate management and R&D by establishing a technological affairs committee.[39] Former Kodak Vice President W.T. Hanson, Jr., outlined the five tasks given the committee:

1. To assess long-term technical opportunities as they emerge from the basic research environment.
2. To establish broad goals for the commitment of Kodak R&D resources. These goals should meet short-term product and process needs as well as long-term technology needs.
3. To ensure that resources are properly allocated to develop the technology necessary to support the longer range business objectives.
4. To approve major corporate product programs including specific goals which encompass the following:
   —schedule,
   —specifications of features and functions,
   —resources required,
   —corporate return, and
   —assessment of risk.
5. To monitor progress in corporate projects and approve any changes which have an impact on corporate goals.

Corporate projects as well as the committee's progress were monitored in weekly meetings, which were chaired by the Chief Executive Officer (CEO). The meetings included top management and the staff directly responsible for projects under review. To Hanson, these sessions represent top management's "real commitment" to R&D planning, an idea that was almost unheard of during Mees' tenure.

Kodak's decision to establish a technical advisory committee was not unique. DuPont and Monsanto, two major chemical manufacturers that are now entering the field of life sciences, increased the power of already existing committees in order to streamline the R&D budget allocation process. Prior to 1979, DuPont had established an executive committee to oversee the activities of the R&D divisions. According to Robert C. Fortney, Executive Vice President of R&D, "the

[33] Ibid, p. 154.

[34] Kantrow, op. cit., does believe that managers' awareness "of the need to **Incorporate technological** issues within strategic decision making" (p. 6) grew during the 1970s. However most of his article emphasized management's refusal to see the connection between technology and strategy, Kantrow (personal communication, 1985) said that *he* believed **corporate managers' views of planning technological innovation have evolved over the past three decades** (as opposed to swinging from one extreme to another).

[35] Edward D Weil and Robert R. Cangemi, "Linking Long-Range Research to Strategic Planning," *Research Management,* May-June **1983, p 33**

[36] Ibid

[37] Ibid.

[38] Ibid.

[39] W. T. Hanson, Jr "Planning R&D at Eastman Kodak *Research Management,* July 1978, p 24

head of central research and each of the individual operating departments had a liaison arrangement with different members of the executive committee who loosely kept track of the whole thing." In 1979, the company was reorganized, and the loose arrangement between the various divisions and top management was replaced by one that was more structured. Fortney believes this change led to "greater corporate involvement in deciding how much of [the total budget] would be in one field, and how much of it would be in another field. "[40] This increase in corporate involvement has forced the members of the executive committee to become more informed. For example, Fortney meets bimonthly with the 11 research directors in the company, reads quarterly research progress reports from the operating departments and the engineering department, reviews monthly reports from the central research and development department, and listens to presentations from the first line people on the results of their research, usually two or three times a month. In addition, he talks informally with the research directors about their budget plans .41 Through these formal and informal meetings, which enable corporate management and R&D managers to share information, DuPont has tried to include technological issues in its corporate strategy.

Like Fortney, Howard Schneiderman of Monsanto strengthened an underutilized committee structure when he became the Senior Vice President for R&D in 1979. Three committees are now involved in the budget allocation process. The technology advisory council, chaired by Schneider-man, is composed of the directors and general managers of R&D and technology. Most of the council meetings focus on the administrative concerns that effect the management of the scientific enterprise. The council is also a forum in which proposals for new or continuing research programs are suggested and discussed. Then, the technology review committee, which is also chaired by Schneiderman and which includes the company's senior executives, evaluates these programs by determining whether or not they are commensurate with the corporation's goals. Finally, the programs' budgets are evaluated and then approved or rejected by the executive management committee, which is composed of top management (including Schneiderman) and which is chaired by the company's CEO. Schneiderman contends that the change in Monsanto's view of the importance of R&D (which is reflected in the change in the budget allocation process) came about as a result of the company's decision to shift its emphasis from producing industrial chemicals to the field of life sciences:

> One way to put it is that Monsanto is now into more brain-intensive and less raw material and capital-intensive businesses than we have been before. And that has some enormous consequences for the way the corporation thinks about research. That is why . . . R&D suddenly moves forward in the corporation's thinking.
>
> In the case of, say commodity chemicals, polystyrene, you don't spend an enormous percentage of your sales on research. You have to spend a reasonable amount, but not 5 or 6 percent. Certainly not 10 percent! An awful lot of money goes into building a plant, and you're spending a lot on your cement in the ground. So the big decisions are capital decisions.
>
> Now the really big decisions are R&D decisions. You're going to see this not in Monsanto alone, but in other companies too. '2

The use of committees and other formal means of communication at Monsanto, DuPont, and Kodak is not at all coincidental. What Schneiderman, Fortney, Hanson and others have learned is that R&D budgeting is "an information and communication process. " While it is true that each company allocates its resources differently, several generalizations about the budget process can be made. After reviewing the literature from the 1970s, Winkofsky, et al., formulated and substantiated these observations:

- the processes are multiperson, involving many persons throughout the organizational hierarchy;
- the processes are multilevel, involving organizational entities at different hierarchies;

[40] Michael F. Wolff, "An Interview With Robert C. Fortney," *Research Management,* January/ February 1984, p. 16.

[41] Ibid., p. 17.

[42] David Webber, "Chief Scientist Schneiderman Monsanto's Love Affair With R& D," *Chemical & Engineering News,* Dec. 24, 1984, p. 11.

- the processes are iterative;
- proposals are passed upward through the hierarchy;
- resource allocations are passed downward through the hierarchy;
- goals in implicit or explicit forms are passed downward through the hierarchy;
- the processes are multicriterion in nature;
- there is no unique set of criteria used by all firms;
- areas of concern which are considered by many firms include: R&D costs and the probability of technical success, manufacturing costs and the probability of commercial success, market potential and the probability of market success, and contribution to corporate goals;
- different levels of the hierarchy may use different evaluative criteria; and
- the R&D budgeting process may be periodic, continuous, or periodic-continuous. [43]

This list contains no mention of the use of financial models. Winkofsky, et al., believe that the complexity of the allocation process "often confounds the formulation of mathematical models." Because different levels of the hierarchy may use different evaluative criteria, there is no firm base on which a financial model can be built and utilized effectively by all parties involved in the decisionmaking process.

Managers and scholars now approach budgeting as a process that relies on shared information and communication for its success; thus, it cannot be easily reduced to mathematical formulae. Despite the plethora of financial and technological forecasting models that have been introduced in the last 15 years, managers have been reluctant to replace qualitative measures with strictly quantitative ones. Most managers use a combination of qualitative and quantitative techniques, depending on the stage of research of the project. Models are used mainly to explore policy alternatives. Qualitative evaluation techniques work best at the level of basic research. A mixture of quantitative and semi-quantitative techniques works well at the level of applied research. Reliance on strictly quantitative techniques greatly increases when a project enters the last stages of product development.

Many executives shy away from using formal analytical methods because the data generated by these models often do not reflect assumptions that are shared throughout the organization. For example, Robert L. Bergen, Jr,, manager of corporate R&D at Uniroyal, is "leery about individuals becoming so committed to a number that they will find it hard to reassess a project later on, adding that there is always a problem when the boss ranks things one way and a subordinate ranks things another. "[44]

The degree to which executives and their subordinates' views are commensurate reflects the level of communication between individuals and divisions within the corporation. Most of the literature in the past 5 years points to gaps in communication as the most important difficulty that has to be overcome if strategy and technology are to be linked. New corporate efforts to mesh strategy and technology are occurring at a time when scientific and technological information is growing exponentially. Management wants information from corporate scientists and engineers and to share this information with representatives from the marketing and manufacturing divisions as often as possible.

[43] Winkofsky, et al., op. cit., pp. 185-187.

[44] Michael F. Wolff, "Selecting R&D Projects at Uniroyal, *Research Management*, November 1980, p. 8

Chapter 5

Government Decisionmaking

Chapter 5

Government Decisionmaking

To be useful to Federal policymakers, quantitative methods for evaluating research and development (R&D) must provide reliable results and fit with existing decisionmaking procedures. As we have seen, neither the economic rate-of-return models nor the noneconomic science indicators can answer all the questions facing policymakers. The economic models do not even meet the needs of industrial private managers, for whom economic payoff is the primary concern. One should not be surprised, therefore, that these models offer little help in making Federal R&D decisions, in which economic payoff is only one of many criteria and often a secondary consideration. In addition, the users of Federal R&D are not captive. Information produced through federally funded R&D is, in most cases, available to anyone who seeks access. Thus, the benefits are dispersed in a way that makes accounting for them nearly impossible.

The goal of federally funded research is not profitability, but a means of achieving social objectives, whether they be health, national security, or the enhancement of knowledge and education. The Federal research infrastructure is designed to provide a stable environment for these goals, despite a changing political environment. This creates an R&D management environment very different from industry, where reorganization is more easily achieved.

In addition, Federal research programs must be responsive to many more groups than industrial research efforts, and this affects the manner by which the research agenda is shaped. The process of obtaining funds from the taxpayer for mission research is complex and quite unlike the R&D decisionmaking apparatus found in industry. The budget process is the first of many hurdles, followed by levels of decisionmaking at the agency level, institute *or* directorate level, program level, and advisory board level. On occasion, Congress has attempted to influence the administration and execution of research programs by using mechanisms such as appropriations riders. Thus, research funding decisions in the Federal Government are subject to levels of review and requirements for accountability unheard of in industry.

To understand the limited utility of quantitative methods for measuring return on Federal R&D, it is important to recognize the complexity of the processes leading to the actual investment decisions. Attempts at evaluating research decisionmaking should be analyzed in the context of the scale and structure of scientific activity in the United States. It is estimated that the Federal Government will be responsible for 49 percent of national R&D expenditures in 1986, up from 46 percent in 1982.[1] The structure of support is pluralistic and decentralized, with 10 agencies responsible for R&D functions. The budgets of each agency differ enormously, as depicted in table 11. Methods for project selection and program evaluation also differ between agencies, reflecting the decentralized and pluralistic nature of the system. These differences are attributable to the age of the agency, the size of the budget, the levels of basic and applied research, agency mission, and the management "traditions" institutionalized over time. In all cases, decisions are made incrementally.

To understand how quantitative methods can be used in Federal decisionmaking, we have to look at the types of decisions that must be made and how these decisions are being made now. Policymakers must establish priorities among all government programs, among the various scientific disciplines, and among projects within a discipline. These priorities are then applied to decisions about research budgets, project selection and termination, and program evaluation. We will look at how these decisions are now made and evaluate the potential for using quantitative methods to assist in the process.

[1] National Science Foundation, Division of Science Resources Studies, *Science and Technology Data Book* (Washington DC: NSF, 1986).

Table 11.- Federal Obligations for Research and Development by Character of Work and R&D Plant: Fiscal Years 1984=85 (thousands of dollars)

Fiscal year and agency	Total R&D and R&D plant	Total R&D	Research: Basic research	Research: Applied research	Development	R&D plant
Fiscal year 1984 (estimated):						
Total, all agencies	46,554,924	44,835,777	6,981,031	8,127,270	29,727,478	1,719,145
Department of Agriculture	925,364	871,942	386,442	455,594	29,906	53,422
Department of Commerce	367,252	360,021	20,522	272,644	66,855	7,231
Department of Defense	27,987,145	27,540,045	816,590	2,168,184	24,555,271	447,1 co
Department of Energy[a]	5,770,604	4,825,576	841,671	1,231,733	2,752,172	945,028
Department of Health and 'Human Services[b]	4,921,924	4,864,292	2,793,052	1,705,911	365,329	57,632
Department of the Interior	427,558	421,825	124,667	276,330	20,828	5,731
Department of Transportation	538,429	515,929	600	81,990	433,339	22,500
National Aeronautics and Space Administration	3,044,400	2,888,900	689,133	1,012,031	1,187,738	155,500
National Science Foundation	1,247,580	1,238,480	1,172,466	66,014		9,100
Veterans Administration	228,100	220,900	15,200	189,700	16,000	7,200
Other agencies	1,096,568	1,087,867	120,688	667,139	300,040	8.701
Fiscal year 1985 (estimated):						
Total, all agencies	54,072,393	52,253,607	7,637,587	8,396,633	36,21 '3,387	1,818,786
Department of Agriculture	926,711	898,941	419,727	449,981	29,233	27,770
Department of Commerce	282,357	270,559	18,416	201,187	50,956	11,798
Department of Defense	34,510,984	34,142,084	913,195	2,408,204	30,822.685	368,900
Department of Energy[a]	6,146,700	4,962,272	944,517	1,268,964	2,748,791	1,184,428
Department of Health and Human Services[b]	4,967,872	4,953,972	2,925,916	1,679,147	348, ! 09	13,900
Department of the Interior	369,209	368,989	102,762	248,556	17, [71	220
Department of Transportation	505,704	495,204	400	79,630	415,174	10,500
National Aeronautics and Space Administration	3,499,400	3,339,400	826,721	1,088,063	1 ,424,(16	160,000
National Science Foundation	1,426,567	1,414,017	1,335,809	78,208		12,550
Veterans Administration	207,600	194,500	15,000	160,000	19,500	13,100
Other Services	1,229,289	1,213,669	135,124	736,693	341,852	15.620

[a]Data shown for fiscal years 1956.73 and fiscal years 1974-76 represent obligations of the Atomic Energy Commission (A EC) and the Energy Research and Development Administration, respectively.
[b]Data shown for fiscal years 1955-78 represent obligations of the Department of Health, Education, and Welfare.

SOURCE: National Science Foundation.

THE R&D BUDGETARY PROCESS

The process of budgeting is a process of compromising among competing values over how funds should be expended. Since funding is essential for any public policy, the budget process deals directly with how values are allocated in a political system. " It is a political process.[2]

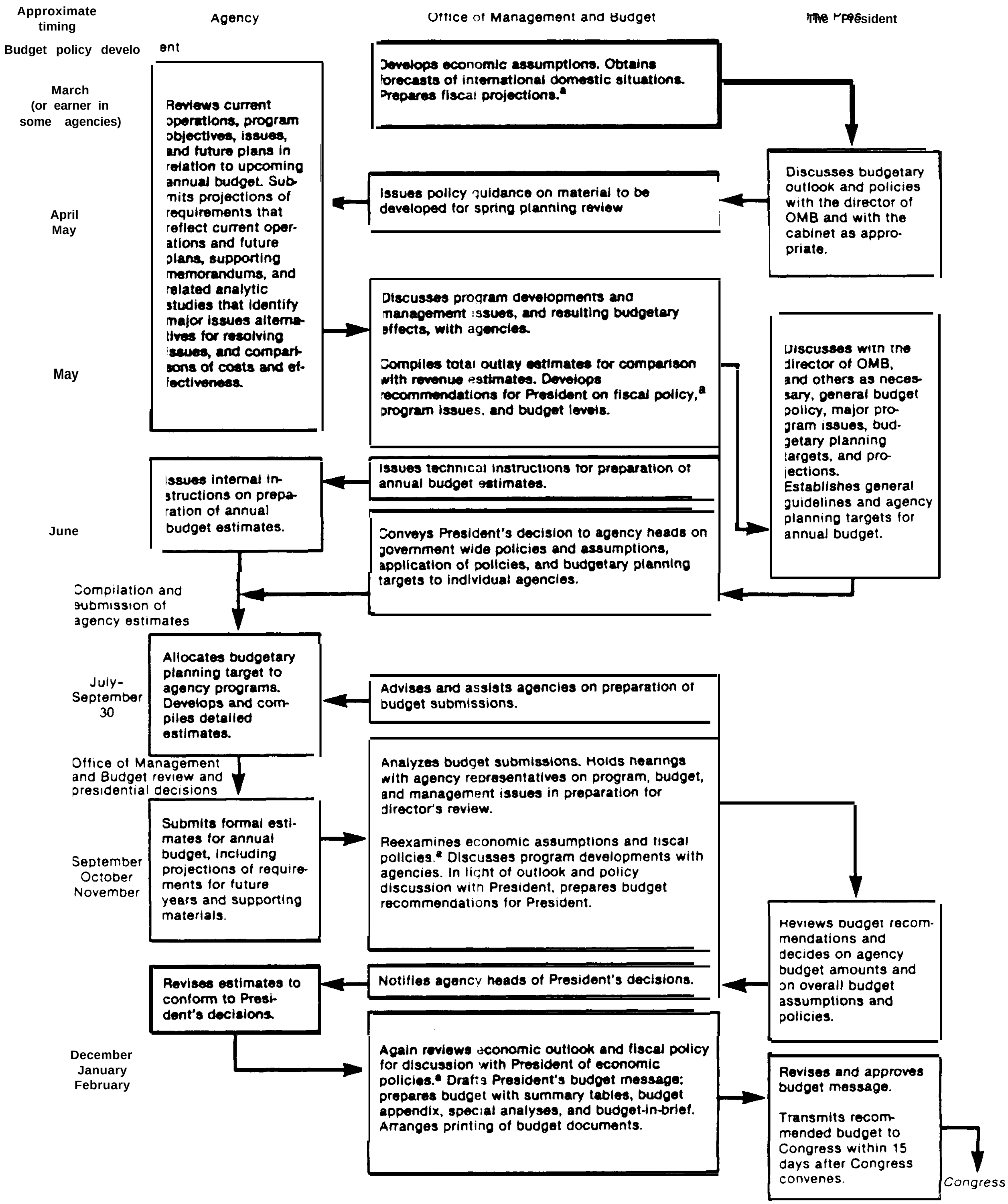
Most descriptions of the Federal budget process include schematic diagrams (see figures 3 and 4) that show the timetable for executive and legislative action. These outlines usually highlight the deadlines for agency budget estimates and the passage of resolutions. The truly significant characteristics of the budgetary process, however, are obscured by the arrows and dotted lines. The process is too complex to be characterized solely by a list of important dates. Outlines and diagrams do not explain how Congress, the Congressional Budget Office (CBO), the President, the Office of Management and Budget (OMB), and the executive agencies formulate recommendations for appropriations and budget outlays. No drawing can adequately represent the influence of the incremental method—the major method for calculating budgets at the Federal level.[3]

Incrementalism informs all aspects of decision-making in normal budget years. Last year's budget is the single most important factor in determining this year's budget, which is the single most important factor in determining this year's author-

[2]William L. Morrow, *Public Administration: Politics, Policy and the Political System* (New York: Random House, 1980), p. 309.

[3]Aaron Wildavsky, *The Politics of the Budgetary Process* (Boston, MA: Little, Brown & Co,, 1984), p 13

Figure 3.—Formulation of the Pesident's Budget



[a]In cooperation with the Treasury Department and the Council of Economic Advisers.

SOURCE William L Morrow, *Public Administration: Politics, Policy and the Political system* (New York Random House. 1980), pp 27-28

Figure 4.—The Congressional Budget Process

Phase	Month	Action	Related action
Congressional action on spending bills	June		
	July	To extent practicable, CBO cost analyses and 5-year projections will accompany all reported public bills, except appropriation bills [sec. 403] Reports on new budget authority and tax expenditure bills must contain comparisons with first concurrent resolution and 5-year projections [sec. 308(a)] If a committee reports new entitlement legislation that exceeds appropriate allocation in latest concurrent resolution, it shall be referred to the appropriations committee with instructions to report its recommendations within 15 days [sec. 401(b)(2)]	CBO issues periodic scorekeeping reports comparing congressional action with first concurrent resolution [sec. 308(b)]
	August		Budget committees prepare second concurrent resolution and report
Adoption of second concurrent resolution and reconciliation	September	Seventh day after Labor Day: Congress completes action on all budget and spending authority bills [sec. 309] 15: Congress completes action on second concurrent resolution [sec. 301(a), (b)] *Thereafter, neither house may consider any bill, amendment, or conference report that would result in an increase over budget outlay or authority figures or a reduction in revenue level adopted in second concurrent resolution* [sec. 311(a)] 25: Congress completes action on reconciliation bill or resolution [sec. 310(c)–(e)] *Congress may not adjourn until it completes action on second concurrent resolution—and reconciliation, if any* [sec. 310(f)]	
	October	Fiscal year begins	

SOURCE; William L. Morrow, *Public Administration: Politics, Policy and the Political System* (New York: Random House, 1980), pp 40, 43

Phase	Month	Action	Related action
Information gathering, analysis, and preparation of first concurrent resolution	October	1: Fiscal year begins [sec. 501]	
	November	10: President submits current services budget [sec. 605(a)]	CBO five-year projections (as soon as possible after October 1) [sec. 308(c)]
	December	31: Joint Economic Committee reports analysis of current services budget to budget committees [sec. 605(b)]	
	January	Approximately last week of month: President submits budget (15 days after Congress convenes) [sec. 601]	
	February		
	March	15: All committees and joint committees submit estimates and views to budget committees [sec. 301(c)] Legislation providing contract or borrowing authority must be made subject to amounts provided in appropriation acts [sec. 401(a)]	Budget committees hold hearings; begin work on first concurrent resolution [sec. 301(d)]
Adoption of first concurrent resolution	April	1: CBO report to budget committees [sec. 202(f)] 1-15: Budget committees report first concurrent resolution (on or before April 15) [sec. 301(d)]	House and Senate consider first concurrent resolution [sec. 305] Conference action and adoption of conference report [sec. 305] Conference report accompanied by joint explanatory statement, which allocates total levels of budget authority and outlays among committees [sec. 302(b)]
	May	Congress—until seventh day after Labor Day—enacts appropriations and spending bills 15: Congress completes action on first concurrent resolution [sec. 301(a)] 15: Deadline for committees to report authorization bills (some exceptions, and waiver procedure) [sec. 402(a)–(e)]	Before adoption of concurrent resolution, neither house may consider new budget authority or spending authority bills, revenue changes, or debt limit changes (some exceptions, and waiver procedure) [sec. 303(a)–(c)] Before reporting first regular appropriation bill, House Appropriations Committee, to extent practicable, marks up all regular appropriation bills and submits to House summary report comparing proposed outlays and budget authority levels with first concurrent resolution [sec. 307] After adoption of first concurrent resolution, each committee subdivides its allocation among its subcommittees, and promptly reports such subdivisions to its house [sec. 302(b)]

SOURCE: House Budget Committee Section numbers are from the Congressional Budget and Impoundment Control Act of 1974

The number of actors involved in the setting of an agency budget is enormous. The three tracks of the budgetary process—authorization, appropriation, and reconciliation-each focus on different dimensions of the agency budget, and each yields its own version. Disparity may occur between the three versions within one Chamber of Congress, as well as between the House and Senate versions. Throughout the congressional budget process, industrial research organizations, scientists, and scientific and professional societies, whose members benefit from research funds, lobby Congress to support increased funding for those programs.

The Federal budget process, while susceptible to confusion and manipulation, is a legitimate attempt to foster some kind of consensus between legislative and administrative budget actors and between competing national policies. In addition, the budget process has become a multi-purpose vehicle for political and policy statements that are not necessarily related to the agency's mission directly. The process is a cobweb of interaction rather than a linear progression from investment to output.

PROJECT SELECTION

Once an agency receives its budget, project selection procedures and styles differ between agencies. In National Science Foundation (NSF) and National Institutes of Health (NIH) investigator-initiated basic research grant programs, the ideas for new projects come largely from the scientific community through the grant application process, consensus conferences, and workshops. The National Science Board and the study sections of NSF, along with the Advisory Councils of the 11 National Institutes of Health, provide additional overall guidance on agency and program direction. There is a high degree of confidence in the peer review process at these agencies and in the scientific community. Questions about the perils of peer review persist but there have been no proposals for change convincing enough to overhaul the system. Until a reliable replacement is found, qualitative, judgmental approaches will dominate in the selection of basic research approaches.

This qualitative approach to project selection has been standard in many agencies since Vannevar Bush recommended it in his 1945 report, *The Endless Frontier.* [7] The first R&D agency to truly implement the concept of peer review as the method for investing in basic research was the Office of Naval Research (ONR), a research agency of the Department of Defense (DOD). ONR uses a peer review process that relies on both in-house and external review, The old ONR model of separating the mission of basic science from the practical mission of the agency provided the model for peer review at NSF.

In comparison, mission-oriented programs in agencies such as DOD, the Department of Energy (DOE), the National Aeronautics and Space Administration (NASA), the U.S. Department of Agricultural, EPA, and agencies with a relatively smaller R&D function, such as the Department of the Interior, the Department of Commerce, the Nuclear Regulatory Commission (NRC), and the Veterans Administration, contract for applied research in support of their technology development or industry support activities, They tend to receive ideas for new projects from a wide variety of sources: industry, Congress, their own program staffs, the national laboratories, and the scientific community. Regardless of the source of a new idea, agency staff will usually conduct a feasibility study to determine whether the new concept is likely to meet cost, performance, and user-acceptability criteria. If the results of the study are promising, the program manager will propose the project as a line item in the new fiscal year budget. The administrative officer in charge of the program area (often an assistant secretary), the head of the agency, the examiners at OMB, and (if it represents a sizable fraction of the program

[7] Vannevar Bush, *Science–The Endless Frontier* a report to the President on a Program for Postwar Scientific Research (Washington, DC: National Science Foundation, 19801 (reprinted from Office of Scientific Research and Development, 1945).

izations and appropriations. Many items in the budget are simply reenacted every year unless there is a special reason to challenge them. In addition, long-range commitments have often been made and the current year's share for previous commitments must be taken out of the total and included as part of the annual budget.[4] These commitments preclude comprehensive assessments of any agency's budget. Thus, actors in the budget process are concerned with relatively small increments to an existing base. Their attention is focused on a small number of often politically controversial items over which the budget battle is fought. Understanding the nature of these battles is crucial to comprehending the entire process.

The inherently incremental nature of the budgetary process precludes in-depth, systematic reviews of programs and agencies. In the past 20 years, two major attempts have been made to infuse some "rationality" into the process, to "change the rules of the game." In 1965, President Johnson announced the implementation of the Planning, Programming, and Budgeting System (PPB) at the Federal level. PPB provided decisionmakers with data from systems analysis, cost-benefit analysis, program budgeting, and cost-effectiveness studies to support decisions about alternative courses of action. However, all of the data generated did not enable policymakers to establish priorities in a more systematic fashion; the program failed to account for the political nature of the budgeting process:

> PPB . . . could determine, within a reasonable margin of error, what the results would be if money was spent for x instead of y. It could also project how much of x and how much of y could be purchased or developed for a specified amount of money. What it could not do was to determine whether it was best to allocate funds for either program x or y.[5]

PPB failed because it set out to tackle an impossible task: the goal of its supporters was to "objectively determine what is inherently ideal, rational, and moral in public policy."[6]

In 1974, Congress enacted the Congressional Budget and Impoundment Act of 1974 (Public Law 93-344) to provide more focus on the "big picture" of the Federal budget. The two standing budget committee and CBO determine the appropriate levels of revenues and public debt for each fiscal year and the subsequent level of total budget outlays and authority. This attempt at budget reconciliation has not affected the incremental nature of preparing separate agency budgets or appropriations bills.

President Carter offered another plan to free the budgetary process from the constraints of incrementalism. He introduced the concept of zero-based budgeting (ZBB). Although its application is a complicated process, its basic purposes and procedures are relatively simple to comprehend. Agencies are directed to bracket their programs into "decision units." Each of these units is assigned a priority status— i.e., the degree to which it is essential to each agency's operations. A minimum expenditure base is supposed to be established by the agencies to represent "essential" program obligations that, therefore, are safe from budget cuts. Theoretically, this is the zero base, with all unnecessary expenditures eliminated. In practice, the base was often much higher than zero. Agencies had a vested interest in protecting certain programs. By increasing the level of the base, the agencies effectively decreased ZBBs effectiveness in evaluating programs. This new, arbitrarily established base only served to increase the budget officials' dependence on the incremental method. Less than 2 years after it was introduced, ZBB was abandoned by all Federal agencies, with the exception of the Environmental Protection Agency (EPA), which still employs this method today.

Incrementalism in the budgetary process is only one difficulty encountered in attempting systematic review of research agency programs. Comprehensive review of Federal R&D efforts is further complicated by the fact that the Federal Government does not have a separate R&D budget. Federal funding for R&D is the sum of those program requests submitted by individual agencies to OMB, subsequently by the President to Congress, and approved, rejected, or altered during the budget review and appropriation process.

---

[4] Ibid.
[5] Morrow, op. cit., p. 309
[6] Ibid., p. 310.

budget) the staffs of the authorizing and appropriating committees in Congress will review the proposal. The project selection process is complete only after a project has been formally included in a congressionally approved budget.

Within DOD, the Defense Advanced Research Projects Agency (DARPA) uses no formal system of peer review for project selection, but relies on the management of contractors by DARPA program managers. Occasionally, the Defense Science Board examines an area of research supported by DARPA in order to make recommendations for future action.[8]

DOE supports R&D carried out through its own laboratories and by contracting with universities and industry. Research is evaluated for funding through the use of peer review, programmatic technical review, and programmatic management review.

NASA relies heavily on internal and external advisory committees for planning future missions, assigning priorities among them, and selecting specific experiments. The NASA Office of Aeronautics and Space Technology research programs select projects by collaborative-review by a network of both researchers and users of research results.[9]

Whether project selection is conducted by peers or agency management, the traditional criteria for selection are based on qualitative judgments. Like industry, government seldom uses quantitative project selection models. A 1974 review of the use of quantitative methods for project selection in government agencies revealed that they were used very little except in a few decisions involving large development projects. Recent surveys reveal no evidence that the patterns and extent of use of quantitative techniques to evaluate proposed research in the Federal Government are undergoing any significant change. *O Recent DOE surveys reached similar conclusions.[11]

J. David Roessner's[12] limited study of the use of R&D project selection models in DOE offers insight into the different ways models can be used in Federal agencies and reveals some of their limitations. He reports that the most extensive use of such models took place in DOE's energy conservation programs. Program managers at the Energy Research and Development Administration and later at DOE were under unusual pressure to justify public expenditures for energy conservation. At the same time, the conservation program was confronted with mountains of proposals that had to be screened and acted on in some systematic, defensible fashion. The models helped screen projects and select those most likely to pay off. They were also used to justify the program to Congress, where models received considerable attention. Congress cared less about the project-by-Project scores than about the aggregate benefit . scores achieved by all projects funded by a program. Quantitative estimates of oil displacement and energy savings generated by the models proved extremely useful in defending expenditures for R&D programs. One successful model compared cost-benefit ratios at the program level based on oil savings with and without Federal expenditures. These models evolved gradually and were applied over several years to project screening and budget decisions. DOE usually used the models with applied research programs, in which the links to commercial applications were apparent.[13]

The Department of Energy's use of cost-benefit analysis for its conservation program is not representative, The connection between technical improvements and economic benefits in energy equipment is much more straightforward than for other technological advances, making a predictive model somewhat useful. Also, economic benefits were the primary goal of the program. Such conditions are rare in Federal research programs, and it is instructive that even DOE has limited its use to a few applied research programs.

'J.M. Logsdon and C.B. Rubin, ***Federal Research Evacuation Activities*** **(Cambridge,** MA: **Abt Associates,** 1985), p. ***14.***

'Ibid., p. 17.

[10]H. Lambright and H. Sterling, ***National Laboratories and Research Evaluation*** (Cambridge, MA: Abt Associates, 1985).

"K. G. Feller, *A* Review *of* Methods *for Evacuating R&D* (Livermore, CA. Lawrence Livermore Laboratory, **1980); and J. David** Roessner, *R&D Project Selection Models in the U.S. Department of Energy* (Atlanta, GA: Georgia Institute of Technology, 1981),

''Roessner, op. cit.

'Ibid., pp. IV-1, 2; V-2.

RESEARCH PROGRAM EVALUATION

Quantitative methods are only occasionally used in evaluating the productivity and relevance of existing programs. A number of agencies are experimenting with quantitative techniques, but few have adopted them for use in systematic evaluation of research programs. Several recent surveys of research managers in the agencies and national laboratories found little evidence of quantitative techniques in research evaluation.

NIH is the only agency consistently using quantitative evaluation methods for accountability and program planning. Its bibliometrics analysis effort is described in chapter 3. The Alcohol, Drug Abuse, and Mental Health Administration is planning to use bibliometric approaches in evaluating research programs in the future. For other agencies and other techniques, we find a history of disappointment.

The National Bureau of Standards (NBS), for example, made an abortive attempt to measure the economic impact of individual projects in the Semiconductor Technology Program. NBS asked firms that subscribed to a program publication to estimate the benefits of each project to the firm and the costs of implementing technical information received from NBS. Agency analysts then compared the social costs with the estimated social benefits. They also used a production function to measure the productivity of the entire program. The objective was to estimate the changes in a firm's productivity attributable to changes in the stock of R&D capital generated by NBS. NBS staff reports that these studies were discontinued because of serious theoretical and methodological problems.[14] NASA's macroeconomic and macroeconomic approaches to measuring the effects of its R&D programs (see ch. 2) also met with serious criticism, and NASA discontinued its efforts.[15]

The Department of Energy employed an elaborate, quantitative evaluation scheme based on peer review to evaluate its Basic Energy Science Program in the early 1980s. Forty small review panels used a formal rating sheet to evaluate 129 randomly selected projects on publications produced, personnel achievements, and project summary descriptions. Panel members rated projects for researcher quality, scientific merit, scientific approach, and productivity. The evaluators compared the results with the scores of comparably funded, nonlaboratory projects also rated by the panels. DOE has not applied this expensive and time-consuming evaluation method to other programs, but the ONR has adopted some aspects of the technique.[16]

In 1982, a National Academy of Sciences panel conducted a study for NSF of approaches to evaluating basic research. They concluded that, beyond peer review, "any additional evaluation procedures should be introduced only if they clearly enhance rather than constrict the environment in which the research proceeds, and that formal techniques cannot usefully replace informed technical judgment."[17] A 1984 survey of 41 research managers in 11 Federal agencies found that nonquantitative methods dominated evaluation:

> Some form of peer review was used by almost every Federal agency, both for selecting individual or team research projects and for exercising managerial control over them. Peer review is also the major way that agencies build a case to demonstrate the value of research they support.[18]

No Federal agency "has in place a research evaluation system which appears to move substantially beyond the organized use of "informed technical judgment."[19]

A review of national laboratory evaluation techniques uncovers a similar picture.[20] Most evaluations of laboratory research are relatively unstructured and do not assume major importance among laboratory activities.[21] The complexity of

[14] Logsdon and Rubin, op. cit., p. 34.

[15] Ibid., p. 15; and Henry R. Hertzfeld, "Measuring the Economic Impact of Federal Research and Development Investments in Civilian Space Activities," paper presented to the National Academy of Science Workshop on "the Federal Roie in Research and Development," Nov. 21-22, 1985, pp. 9-12 and 16-21,

[16] Logsdon and Rubin, op. cit., pp. 26-28: and Lambright and Stirling, op. cit., p. 28.

[17] Logsdon and Rubin, op. cit., p, 38,

[18] Ibid., p. 25.

[19] Ibid., p. 38.

[20] Lambright and Stirling, op. cit.

[21] Ibid.

lab roles, which include research performance, research management, and entrepreneurship, precludes most formal, quantitative evaluation techniques.[22] some laboratories occasionally use structured peer review and bibliometric techniques, often performed by outside contractors, but lab managers view these as "supplements to the less structured evaluations, rather than substitutions."[23]

Although economic models for R&D performance came into use in the 1950s, government research managers have not adopted them for program evaluation. This reflects the nature of research manager concerns as well as the accuracy of the models. Research managers are responsible primarily for the quality of research in their programs, and economic payoff does not necessarily reflect the quality of research. A breakthrough in basic or applied government research does not guarantee an economic benefit. No economic effect will occur unless a private company decides to incorporate the breakthrough in a product, and the success of that product depends on such factors as the availability of capital, effective product development, consumer interest, marketing skill, tax and regulatory environment, and competition. Research managers have no say in these other factors and no control over the commercial uses of the research they manage. They therefore limit their attention to what they can control—the quality of research.

Bibliometric methods offer a quantitative measure of research quality, but one not reliable enough to serve as the sole basis of research evaluation. These methods are, however, a useful supplement to informed technical judgment and the peer review process.

'Ibid., p. 9.
'Ibid., p. 10.

FORECASTING AND STRATEGIC PLANNING

The extent of agency use of strategic planning to identify promising future directions for civilian research is not known. DOE's Energy Research Advisory Board, NSF's National Science Board, and NIH's scientific advisory councils provide guidance that might pass for strategic planning. The NRC and its constituent bodies,—the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine—carry out numerous reviews of research programs and research fields for the executive and legislative branches. The most comprehensive of these are the *Research Briefings* and *Five Year Outlooks* prepared by the Committee on Science, Engineering and Public Policy. Since all NRC reports are prepared by committees of scientists, they represent, to some degree, the informed, consensus-based peer judgments of the scientific community on the state of the research enterprise. None of these, however, can be said to constitute true strategic planning or forecasting. For examples of more systematic forecasting and planning activities related to science and technology, one can look to Japan.

Planning for Innovation in Japan

The Japanese Science and Technology Agency (STA), established *by* the Japanese Government in 1956, has relied on strategic planning for its success in identifying technological innovations. In 1969, STA's responsibilities expanded to include funding research as well as coordinating research activities among the various government ministries and agencies. In addition, STA began to make technological forecasts. After the results of the first study were obtained, the agency realized these forecasts would be vital to the creation of rational, long-term research policies and made them a regular part of its operating procedures.

The forecast effort begins by identifying economic and social needs. Forecasters then survey research areas to identify potential scientific and technological developments that can meet these needs. They then establish priorities for various R&D plans. Each forecast is presented in two ways: one that is "exploratoy or predicative, relating to individuals' expectations of change given their accumulated knowledge and experience, and another that is normative—that involves setting an objective and a time-scale within which it is to be achieved. "[24] The forecast is based on a survey of over 2,000 people from government, academia, and industry with a broad knowledge of several scientific and technological research areas, who not only answer questions but who also comment on their colleagues remarks.

Industrialists, academics, and government officials have cited four benefits from the survey process:

1. The studies provide a mechanism to ensure that researchers in all sectors, along with policymakers in government and industry, are periodically forced to think systematically about the long-term future.
2. The forecasts yield a general summary of what is happening, or likely to happen, across the entire range of R&D activities. They therefore permit more "holistic" vision-making, enabling the potential longer term cross-impacts of developments in ons research field on another to be identified at an early stage.
3. By surveying comprehensively the intentions and visions (and thus indirectly the current strategic R&D activity) of the industrial research community, the surveys provide a useful mechanism for synthesizing major research trends across science-based sectors. . . . It is the existence of such surveys that is in part responsible for the strong agreement among Japanese firms as to what are likely to be the critical future developments in their sector.
4. The STA forecasts provide a useful mechanism for helping government establish national priorities in allocating resources. Requirements for infrastructural support can be identified from the "bottom-up" by industry, rather than being imposed by state planners who may not always be in touch with industrial problems. Although such forecasts do not, in themselves, lead directly to policy decisions, the systematic information which they generate helps narrow down the range of different views that can be held on a particular R&D related issue, bringing eventual consensus that much closer. [25]

In addition to the STA, the Ministry of international Trade and Industry (MITI) and the Agency for Industrial Science and Technology (AIST) are responsible for providing funding and long-term guidance for applied research and development. Each MITI division, representing a major industrial sector, establishes a long-term plan, which is revised every 3 to 5 years. Officials from both the ministry and the agency try to incorporate these plans into MITI's vision of the future of Japanese industry. This vision serves as the basis for MITI and AIST's long-term R&D plan, which is also revised every 3 to 5 years. This plan enables the ministry to spot research trends early and *to* predict how new technologies might develop. These predictions, in turn, form the basis for an R&D policy that concentrates on initiating research programs in areas of strategic importance. Like STA's forecasts, MITI's visions are based on suggestions that come from the bottom up. Irvine and Martin stress that the myth of "Japan, Inc. " has arisen because foreign commentators have often overlooked this point:

> The process involved here is not one of centralized "top down" planning by MITI, which then imposes its objectives on industry and others (who do not question them), Instead, most influence tends to flow in precisely the opposite direction, with MITI's role largely confined to "tapping into" the views and firms where consensus lies. Only as a last resort are priorities imposed—for example, to give *one* industrial sector's agreed program precedence over another's.[26]

Before the government introduces a new policy members of an informal working group, which represents a major trade association, meet to discuss common long-range goals. These member: are employees of firms that continuously monitor R&D developments throughout the world They also have access to the company's interna forecasts of technological innovations. The in formed comments of these members are used by consulting groups when they design surveys of in

"John Irvine and Benjamin R. Martin, ***Foresight m Science: Picking the Winners*** (London: Frances Pinter, 1984), p. 108.

"lb[d., pp. 110-111.

"Ibid., p. 118.

dividual industrial sectors .27 The working groups' suggestions and the sectorwide surveys provide an essential link between STA's and MITI's macroforecasts and firms' internal forecasts for specific products .28 According to Martin and Irvine:

> From the point of view of industry, the sector forecasts, because they are much more specific than the macroforecasts, are much more valuable for planning corporate R&D strategy. Equally, the sector studies, based as they are on a synthesis of industrial views, constitute a key input into discussions within MITI and STA.[29]

MITI's role is to construct the broad framework within which a consensus on long-range R&D goals can be established; to catalyze the formation of consensus by sorting and publicizing the results within the relevant sectors; and to try to build a consensus among the various industrial sectors as to long-term R&D priorities.

Several lessons can be learned from MITI's consensus-generating approach to strategic research forecasting. First, forecasts that successfully identify research areas of long-term strategic importance are based on up-to-date background information on research trends gathered from industry, academic, and government reports from around the world. Second, the forecasts incorporate "technology-push" and "market-pull" perspectives because scientific and technological advances must be coupled with changing market demands for technological innovations to be successful. Third, there are a number of advantages that can be gained from adopting a bottom-up approach to forecasting rather than a centralized, "top-down" approach.

> Apart from being dependent on a narrow range of information inputs, "top-down" forecasts and the resultant research policies are more likely to antagonize not only the basic science community (which may feel that it has been inadequately consulted in the forecast process), but also industry (which naturally tends to feel that it is in the best position to judge the commercial prospects for strategic research). 'o

"Ibid,, p 301
'® Ibid., p 129
'Ibid., p. 128
'Ibid., p. 143

The last and perhaps most important lesson to be gained from the STA and MITI forecasts is that the process of generating the forecasts is much more important than the product—the specific results they yield. The process unites people from different groups and different professions within those groups and provides a framework within which they can "communicate directly or indirectly (through a Delphi-style forecast) with each other. "[31] Policymakers, professional forecasters, scientific analysts, and academic and industrial researchers are periodically forced to think about long-term R&D activities by the process. This enables them to coordinate research plans and to form a consensus on priorities for future strategic research. Furthermore, the process generates a feeling of commitment to the outcomes of the forecasting studies. Thus, the predictions become self-fulfilling prophecies. The Japanese contend that these five C's—communication, *concentration on the future, coordination, creation of consensus, and commitment—have* benefited their strategic planning efforts tremendously. Until now, these benefits have outweighed such disadvantages as forecasts' tendency to encourage conservativism and breed excessive competition. Martin and Irvine warn that this balance might be upset in the future "as the Japanese place increasing emphasis on more basic research (where creativity and unconventional approaches are clearly at a premium)."[32]

While other students of Japan warn against placing too much faith in the apparent tidiness and completeness of the framework building process, the importance of wide participation in goal-setting is apparent.

The centrally coordinated Japanese R&D system has served Japan well in applying basic research findings. Yet, pluralism in the U.S. R&D system encompasses several attributes. In testimony before the Task Force on Science Policy, Rodney Nichols of Rockefeller University states that the pluralism of the system "hedges against

'Ibid., p. 144.
"Ibid.

fluctuations in the fashions and policies influencing any lines of R&D support. "[33]

A full-blown pluralistic system depends upon a high level of sustained R&D. It runs the risks of some redundancy when sponsors overlap their support—surely at the research end of the spectrum, where there are many, small projects underway through many sponsors. By doing so, it gains the long-run advantages of giving all missions a window on research.

Pluralism also protects against the inherent frailty, even occasional ignorance, of decisions by research managers. It aims, in principle, to strengthen the broad swath of R&D by being aware of how unpredictable are the origins of great ideas: and how unpredictable are the consequences of results that first seem mere curiosities. Thus, some funds go to all good ideas in order to ensure that the few seen later to be the best have had a chance.[34]

"Rodney W. Nichols, testimony presented before the U.S. House of Representatives, Committee on Science and Technology, Task Force on Science Policy, Oct. 23, 1985.

"Ibid, pp. 10-11.