NIST GCR 03–857

A Toolkit for Evaluating Public R&D Investment

Models, Methods, and Findings from ATP's First Decade

Prepared for
Economic Assessment Office
Advanced Technology Program
National Institute of Standards and Technology
Gaithersburg, MD 20899-4710

By
Rosalie Ruegg
Managing Director
TIA Consulting, Inc.
ruegg@ec.rr.com

Irwin Feller
Senior Visiting Scientist
American Association for the Advancement of Science
Professor Emeritus
Pennsylvania State University
iqf@psu.edu

July 2003

U.S. DEPARTMENT OF COMMERCE
Donald L. Evans, Secretary

TECHNOLOGY ADMINISTRATION
Philip J. Bond, Under Secretary of Commerce for Technology

NATONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Arden L. Bement, Jr., Director

Contents

PART I: EVALUATION FRAMEWORK

PART II: APPLYING EVALUATION METHODS TO ATP

Figures

Tables

Note from Project Manager

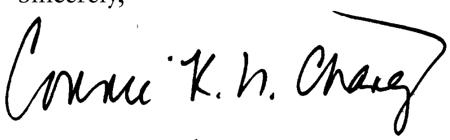
The seed for the idea of constructing a toolkit of evaluation tools for program managers came from attending a workshop I co-organized on "Financing and Managing Technology Development, in Veracruz, Mexico in April 2001. This workshop was held during Mexico's 13th International Congress for Science and Technology and brought together program managers, officials, and consultants from Norway, the European Union, Canada, the United States, and Mexico to share their experiences. Presenters addressed the financing methods and time-frames for public support of technology development; program management, project selection, and oversight; and benchmarking and evaluation of impact.

Rosalie Ruegg, the co-author of this report, gave a presentation at this workshop on evaluation fundamentals that led to a lively exchange over the most effective evaluation tools and methodologies to achieve policy goals. It was during the plane ride home that the idea for a toolkit for evaluating public R&D investment took root.

The timing of the project was perfect because it came on the heels of a National Research Council report (*The Advanced Technology Program: Assessing Outcomes*, 2001) that found that ATP "set a high standard for assessment involving both internal and independent external review." Over the past thirteen years beginning in 1990, the Advanced Technology Program through its Economic Assessment Office has been sponsoring rigorous and groundbreaking scholarship that has advanced the understanding of the process of technology-based innovation.

It is with the hope of sharing what we have learned and to educate those new to program evaluation that we bring to you this toolkit. The authors and I welcome your comments and suggestions.

Sincerely,



Connie K.N. Chang
Supervisory Economist
Leader, Policy Research & Analysis Group
ATP Economic Assessment Office
connie.chang@nist.gov

About the Authors

Rosalie Ruegg, managing director of TIA Consulting, Inc., specializes in the economic assessment of new technologies. Recent clients include the National Academies of Science, Harvard University, the Department of Energy, the State of Indiana, Finland's national technology program, Johns Hopkins University, other agencies of government, and other private and non-profit consulting companies. Prior positions include director of the ATP's Economic Assessment Office, senior economist in NIST's Center for Applied Mathematics, and financial economist for the Federal Reserve System's Board of Governors. Among her more than 60 publications is a case-study guide for science managers, proceedings of an international workshop on evaluation, and an economics textbook. She has served on various editorial boards, most recently as economics editor of Macmillan's *Encyclopedia of Energy.* A former member of the Federal Senior Executive Service, she received the Department of Commerce Gold Medal Award for excellence and the Institute of Industrial Engineers' 2001 Wellington Award for contributions to the field of engineering economics. She received degrees in economics from the Universities of North Carolina (Phi Beta Kappa, cum laude) and Maryland (Woodrow Wilson Fellow), an M.B.A from The American University, and extensive executive training.

Dr. Irwin Feller is Senior Visiting Scientist, American Association for the Advancement of Science and Professor Emeritus of Economics, The Pennsylvania State University, where he served on the faculty from 1963–2002, and as director, Institute for Policy Research and Evaluation from 1977–2002. Dr. Feller's current research interests include the evaluation of federal and state science and technology programs, the economics of academic research, and the university's role in technology-based economic development. He is the author of *Universities and State Governments: A Study in Policy Analysis,* and over 100 refereed journal articles, final research reports, and book chapters. He has been a consultant to the President's Office of Science and Technology Policy, National Aeronautics and Space Administration, the Carnegie Commission on Science, Technology, and Government, the Ford Foundation, the Cleveland Foundation, National Science Foundation, National Institute of Standards and Technology, U.S. General Accounting Office, U.S. Department of Education, the U.S. Department of Energy, and to several state governments and universities. He has served on National Academies committees on international benchmarking of U.S. science and manufacturing modernization, and chaired the American Association for the Advancement of Science's Committee on Science, Engineering, and Public Policy. He chairs the National Science Foundation's Advisory Committee for Social, Behavioral, and Economic Sciences, and is a member of its Advisory Committee on the Government Performance and Results Act, the Institute of Medicine's Committee for Assessment of Excellence Programs at NIH, and the National Research Council's Transportation Research Board's Research and Technology Coordinating Committee.

Acknowledgments

We express a special thanks to Connie Chang, ATP supervisory economist and the project manager, for her meticulous attention to all aspects of the endeavor, inspiration, and valuable comments and advice throughout. We also thank Stephanie Shipp, Director of ATP's Economic Assessment Office; Jeanne Powell, Shawn McKay, and Prasad Gupte, also of the Economic Assessment Office; Barbara Cuthill, Group Leader for Information Technology and Applications, ATP; and John Hewes, former ATP program manager, for their careful reviews and many helpful comments. Elissa Sobolewski, Acting ATP Deputy Director, and Brian Belanger, former ATP Deputy Director, reviewed the final version of this report. We acknowledge the central role played by the many evaluators who conducted the body of 45 evaluation studies from which our report draws. We also acknowledge the excellent editorial assistance provided by Sheila Turnage, technical writer.

Abstract

Evaluation is an essential component of publicly funded R&D programs, both in support of program management and public policy. The Advanced Technology Program (ATP) has emerged over its first decade as a leader in evaluation, engaging nationally prominent evaluators to apply new and existing methods in building an analytical and empirical basis for ATP's operations and performance.

This report draws from a body of 45 studies commissioned by ATP between 1990 and 2000 and analyzes the methods and techniques used and examines the findings of those studies. These studies have increased understanding not only of ATP but also of the dynamics of innovation systems and the relationships between public and private sector funding of R&D. The findings examined are organized around five major themes: firm/industry effects, collaboration effects, spillover effects, interfaces and comparisons with other programs, and measures of overall program performance.

The extensive toolkit of evaluation methods presented in the report illustrates how those methods can be used to answer a variety of stakeholder questions. Methods include survey, descriptive and economic case study, bibliometrics, historical tracing, econometrics, expert judgment, social network analysis, cost index, and a composite performance rating system constructed from indicator metrics. Additionally, the use of analytical and conceptual modeling to explore a program's underlying relationships and process dynamics is considered. The political economy of ATP is discussed, and an evaluation framework and an overview of evaluation best practices are provided.

The report integrates and condenses a large body of related research and thus provides ATP with a convenient reference work, toolkit, and planning guide. For those administrators of other programs, public policy makers, and evaluators, the report also serves as an evaluation toolkit by providing a logical framework

for program evaluation, illustrating the use of evaluation methods and techniques, providing an overview of evaluation principles and practices, organizing a body of knowledge on how public-private partnership programs function, and contributing to an understanding of what evaluation is and how it is practiced in the field of R&D.

Keywords: Advanced Technology Program, assessment, economic evaluation, evaluation methods, impact analysis, logic models, public policy, public-private partnership program, R&D, spillovers, technology

This research was conducted between July 2001 and December 2002.

Executive Summary

Evaluation is a powerful tool for decision makers, but only if it is correctly structured, managed, and applied. Among federal and state agencies interested in science and technology, the Advanced Technology Program (ATP), located in the U.S. Department of Commerce's National Institute of Standards and Technology (NIST), has emerged as a leader in the effective use and development of evaluative tools. Over its first decade, ATP followed a multi-faceted approach to evaluation, providing a mosaic of findings about how the program works and its impacts.

This report assembles a large body of ATP's evaluation studies into a coherent framework, making the studies more accessible and understandable to a diverse audience. An expected benefit is better utilization of past evaluation and increased efficiency and effectiveness in planning future evaluation. In effect, the report provides an evaluation "toolkit" for ATP that will also be useful to others who operate public technology programs. The toolkit provides an evaluation framework; a directory of evaluation methods, tools, techniques, principles, explanatory information, and best practices; an account of ATP's use of evaluation models and methods over its first decade as revealed in a body of 45 selected studies; a cross-cutting compendium of findings; and recommendations for future work. The report addresses the science, craft, and art of evaluation in the context of ATP. It shows how a program established in a climate of political and conceptual debate can use evaluation techniques to answer questions about its fundamental rationale, design features, and economic impacts.

Part I provides a general framework for evaluation, discussing evaluation fundamentals and methods, best practices, and an evaluation logic model to describe ATP's evaluation program. Part II demonstrates the use of evaluation methods by drawing on ATP evaluation studies. Part III presents the emerging body of knowledge from studies of ATP—knowledge about firm behavior, collaboration, spillover effects, interfaces with state and international technology programs, ATP's performance at large, and knowledge about evaluation itself. Part III also presents the authors' conclusions and recommendations. Other features include a glossary of terms, methods bibliography, and a quick reference guide to evaluation models and methods, ATP studies cited, and study findings on program impacts.

Evaluation Underpinnings

For a public sector program like ATP, evaluation seeks to measure change and to determine if the change is attributable to program intervention. An effective evaluation program should investigate change in terms of a program's mission-driven goals, and should compare its findings against intended results. As a point of departure, this report starts with a generic logic model of program evaluation, depicted below, and fleshes out the model using ATP as illustration.

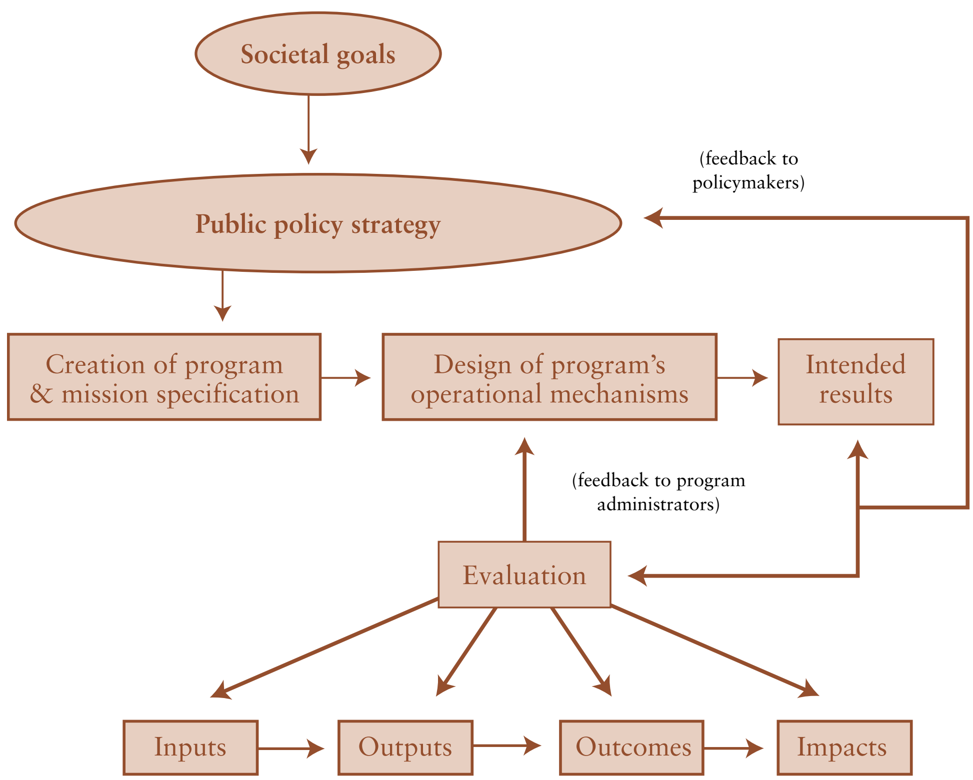
The report summarizes the major analytical themes economists and others use to explain the rationale for ATP: (1) Global economic competition is increasingly driven by technological advance; (2) enabling technologies tend to generate large spillovers; (3) high level of technical risks contribute to an R&D funding gap in the private sector; (4) many advanced technological development projects require multi-disciplinary and multi-organizational collaborative efforts; and (5) the nation's capacity for economic competitiveness and prosperity depends in large part on its innovative capacity, which can be strengthened through public-private partnerships.

ATP's evaluation program has emphasized modeling its underlying program theory—exploring basic concepts, developing underlying causal maps, developing and refining analysis models, and investigating paths connecting program activities to intended impacts. Findings from ATP's studies in turn have shaped its program and evaluation design in numerous ways.

Multi-Faceted Methodological Approach

Evaluators use a variety of evaluative methods, each with its advantages, disadvantages, and specialized purposes. In a multi-faceted approach, like that used by ATP, methods are chosen for their appropriateness to the question at hand, to cost and administrative feasibility, and to a purposeful mixture of methodological

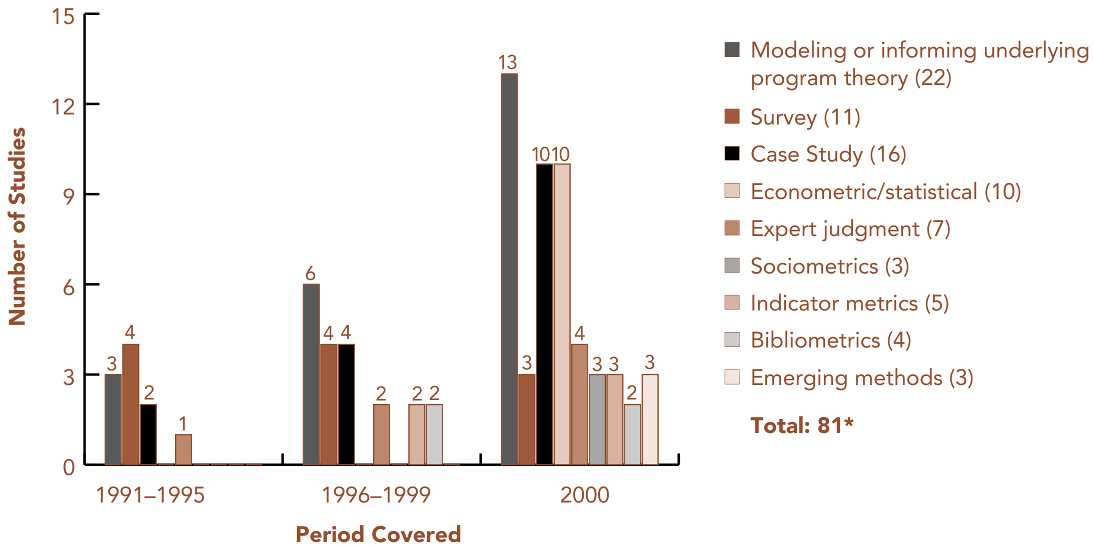
Generic Evaluation Logic Model



Note: The dynamics by which the transformation among these various stages occur are often complex and are themselves the subject of evaluation.

paradigms. Three dominant characteristics of ATP's evaluation program have been the care with which methods and techniques have been matched to the questions being posed, the evolution toward more rigorous tests of causal relationships between ATP activities and observed outcomes, and the development of new tools when existing tools were not up to the task. The result is an extensive and increasingly sophisticated toolkit of methodologies available to evaluate ATP and other technology programs. The figure below depicts the major methods used by ATP over its first decade, and the changing intensity of their use over time. For

ATP's Evolving Use of Methods Over its First Decade



*These 81 methods are employed in the 45 ATP studies commissioned between 1990 and 2000 that are examined in this report.

example, the use of case study increased from 2 to 4 to 10 between the period 1991–1995, 1996–1999, and 2000, respectively.

An Emerging Body of Findings

Throughout its history ATP has had to demonstrate that its operations added to, rather than displaced, the actions of the private sector in assembling the capital necessary to nurture high-risk, enabling technological innovations. It also has had to prove that ATP assistance produces economic benefits that extend beyond the direct recipients of ATP awards to generate broad benefits for the nation.

Evaluation has provided descriptive and analytical information on program recipients and program outputs to ATP and NIST officials, to key executive and congressional decision makers, and to other stakeholders, including the general public. The body of evaluative work conducted over ATP's first decade has answered central questions arising from ATP's mission.

A crosscutting analysis of the evaluation studies reviewed revealed much information that bears directly on ATP's mission-driven goals. This analysis is organized around the following major themes: (1) firm/industry effects, (2) collaboration effects, (3) spillover effects, (4) interfaces and comparisons with other programs, and (5) measures of overall ATP performance, including portfolio analysis, social returns on investment, and impacts on competitiveness. Taken as a body of work, these studies have also contributed to enhanced understanding of the dynamics of the U.S. innovation system, particularly the characteristics of productive R&D relationships between the public and private sectors.

Firm/Industry Effects

Findings on private firm effects, drawn from 13 studies, indicated that ATP substantially expanded and enhanced the R&D activities of the companies examined. The studies provided a growing body of evidence that ATP funding is complementary to, not a substitute for, private sources of R&D funds. They indicated that ATP funding leverages and accelerates R&D, refocuses R&D on more technically challenging problems and enabling platforms of technologies, and fills a significant funding gap. One study concluded that the median time-savings per project was three years and the median economic value to the company per year saved was $5–$6 million. Two other studies estimated significant program-induced increases in patenting by ATP award recipients, indicating a positive impact of ATP on firm research productivity. With regard to the participation of small firms in ATP, the research showed robust participation rates and strong project performance relative to companies of larger size.

Collaboration Effects

The report drew findings on collaboration from 10 studies. One recurring conclusion was that there are high rates of collaboration in ATP projects, including formal joint venture members and extending strongly to single-applicant companies. For example, 84% of the first 50 completed projects entailed collaborative relationships, ranging from R&D partnerships with other firms, universities, and non-profit labs, to alliances with other firms to pursue commercialization. These studies found that ATP successfully encouraged applicants to propose projects entailing collaboration, frequently with entirely new partners. Collaborations of firms with universities was a topic of several of the studies on collaboration. Findings were that collaborations with universities were frequent and that they enhanced the research capabilities of the firms and provided an avenue of knowledge diffusion from and through the universities.

Interestingly, studies also found that the collaborations were frequently fluid, with changes among collaborators occurring during a project's life cycle. Some of these changes may be positive, keeping true to ATP criteria, while others may represent deviations, such as the loss of key participants or a retreat from the more challenging research goals, requiring ATP managerial intervention. The studies suggest that by monitoring projects throughout their lives, ATP is able to respond to and manage change.

According to a study of joint venture participants, ATP contributes to joint venture success by: (1) accelerating the development of high risk technologies, (2) increasing project stability, (3) getting projects through particularly difficult periods in their life cycles, (4) overcoming barriers to collaboration, and (5) increasing up-front planning. Project participants identified specific benefits (particularly a positive effect on creativity), and costs (primarily increased administrative burden) associated with collaboration. Almost all project participants involved in collaborative arrangements indicated that their experience with ATP has stimulated them to plan additional collaborations. Among factors important to the success of collaborative relationships, the study corroborated other work that found establishing an environment of trust to be critical.

Spillover Effects

The concept of economic spillovers occupies a central place in the case for a public sector program like ATP and has helped shape many of ATP's program design features. Findings from 10 of the studies increased understanding of ATP's success in generating spillovers. The studies provided considerable evidence that ATP-funded projects generate outputs—publications, patents, patent citations, collaborative linkages, and products—that will potentially lead to knowledge and market spillovers. The potential of network spillovers was also identified, but not yet measured.

One study concluded that ATP selects projects with attributes conducive to generating large knowledge spillover effects. Those attributes included linkages to other organizations, and a positive attitude of award winners toward information sharing. Several studies concluded that the degree to which a funded company is embedded in organizational networks is a major factor in knowledge spillover potential.

Study results also indicated that ATP selects projects whose firms have more extensive ties to other businesses, and, hence, are better positioned to realize commercial success and related market spillovers. In the studies examined, quantitative estimation of the economic value of spillover benefits was limited to market spillovers. Where estimated, market spillover benefits appeared large, and far in excess of private benefits. Among the body of work examined, none of the studies estimated the economic value of both market spillovers and knowledge spillovers.

Interfaces with State Programs and Comparison with Counterpart Programs Abroad

This report draws on five studies for data on the interactions between ATP and state programs and on ATP-counterpart programs in other countries. One study's major conclusion, based on analysis of existing state technology programs, was that state technology programs span the research and development continuum, but cluster around the downstream applied/commercialization segment rather than the upstream research segment of the continuum. The study found that ATP, in contrast, centers its activities on technical challenges, supporting work primarily in the concept and development phases. A collection of case studies highlighted the possibilities of firms combining support from both ATP and state government programs, and illustrated how ATP and the state programs can augment one another. With regard to counterpart programs in other countries, one study offers a framework for standardizing the comparison of ATP with foreign counterpart programs. This systematic approach has helped ATP meet its mandated requirement to test for eligibility of foreign-owned companies for awards and has allowed ATP to learn from the experience of other programs.

Overall ATP Performance

Thirteen studies provided findings on ATP's impact on national industrial competitiveness and the national capacity to innovate, its ability to deal appropriately with failed projects, its contribution to social benefits, and its overall effectiveness. Prospective case studies provided evidence that the benefits of the program far exceed its costs. These studies collectively attributed to ATP more than $15 billion in expected present value social benefits from just a few projects, much greater than the total amount spent by the program. As expected, not all of the projects are strong performers, but several years after project end an estimated 16% of completed ATP-funded projects showed strong progress toward creating and disseminating knowledge and commercializing projects and processes, and another 26% also showed substantial progress. Five to 6% of all funded projects failed to start or were terminated prior to completion for a variety of reasons. In a major independent assessment, the National Research Council concluded that ATP is effectively meeting its legislative goals.

Recommendations for Future Directions

This report concludes by proposing future directions for ATP's evaluation program, taking into account stakeholder questions, gaps in coverage, past accomplishments, and promising research opportunities. The authors provide 10 recommendations, in no particular rank order, as follows:

- Increase retrospective, market-data–based analyses
- Incorporate both direct- and indirect-path analysis in benefit-cost case study, including estimates of both market and knowledge spillovers
- Continue status reports of completed projects and, on a sample basis, repeat them further out in time
- Update information on state and foreign counterpart programs
- Further develop several of the promising new evaluation techniques
- Deepen analysis of knowledge spillovers beyond patent-only–based studies
- Identify and address new questions that arise as ATP is modified
- Pursue analysis of failures and successes
- Continue an effective mix of in-house and external evaluation studies
- Take greater advantage of evaluation results in decision-making processes

In sum, evaluation has provided an objective analytical and empirical basis for assessing ATP's operations and impacts during its first decade of operations. Cumulatively, these evaluations highlight the value of applying multiple evaluation methods to complex problems, building a body of credible evidence over time that ATP is achieving its objectives.

Main topics covered in the report are highlighted below.

Highlights of Main Topics

MODELS AND METHODS

- ✓ ATP's evaluation logic model
- ✓ Generic treatment of evaluation methods: list, definitions, examples of use
- ✓ Chronological listing of 45 ATP evaluation studies commissioned (1990–2000), with principal and secondary methods used by ATP
- ✓ ATP's use of evaluation methods*
  - Modeling or informing underlying program theory (22 supporting studies covered in the report)
  - Survey method (8 of 11 supporting studies covered in the report)
  - Case study method (10 of 16 supporting studies covered in the report)
  - Econometric/statistical methods (8 of 10 supporting studies covered in the report)
  - Expert judgment method (5 of 7 supporting studies covered in the report)
  - Sociometrics (3 supporting studies covered in the report)
  - Indicator metrics (5 supporting studies covered in the report)
  - Bibliometrics method (3 of 4 supporting studies covered in the report)
  - Emerging methods (3 supporting studies covered in the report)
    - Cost index method
    - Social network analysis/fuzzy logic
    - Composite performance rating system

*Some studies used multiple methods. Not all studies are referenced in each chapter.

continued on next page

Highlights of Main Topics (Cont'd)

CROSSCUTTING FINDINGS

✓ Impact on private firms
- Financing gap
- Halo effect
- Acceleration
- Firm productivity
- Small firm participation
- Commercialization, company growth, and private returns

✓ Collaboration
- Activity, structure, formation, and attribution
- Changes in relationships
- University representation and roles
- Determinants of success
- Benefits and costs

✓ Spillover effects
- Market spillovers
- Knowledge spillovers

✓ State and foreign programs
- State program interfaces
- Foreign program comparisons

✓ Overall ATP performance measures
- ATP's contribution
- Improving competitiveness of the United States and its businesses
- Fostering the national capacity to innovate
- Dealing with failed projects
- Measuring progress, social benefits, and overall effectiveness

Acronyms and Abbreviations

ATP	Advanced Technology Program
BRS	Business Reporting System
CAFE	Corporate average fuel economy
CCAR	Closed-cycle air refrigeration
CPRS	Composite performance rating system
DCCT	Diabetes Control and Complication Trial
DDS	Digital data storage
DSG	Diamond Semiconductor Group
EAO	Economic Assessment Office
FCM	Flow-control machining
GAO	General Accounting Office
GDP	Gross domestic product
GPRA	Government Performance and Results Act of 1993
IIH	Information Infrastructure for Healthcare
IRR	Internal rate of return
MEMS	Micro-electromechanical systems
NAS	National Academy of Sciences
NBER	National Bureau of Economic Research
NSIC	National Storage Industry Consortium
NIST	National Institute of Standards and Technology
NPV	Net present value
NRC	National Research Council
OIG	Office of Inspector General
PDM	Pit depth modulation
PWB	Printed wiring board

QALY	Quality Adjusted Life Year
REMI	Regional Economic Modeling, Inc.
RTI	Research Triangle Institute
STEP	Science, Technology, and Economic Policy
SWAT	Short-wavelength sources for optical recording

Introduction

Evaluation is a powerful tool for policy and decision making, but only if it is correctly structured, managed, and applied. At the technical level, this means having a clear set of objectives, a logical framework, and valid methods and findings. At the program level, this means having knowledgeable evaluators and program administrators who can work with evaluators to structure relevant questions, relate general findings to specific agency settings, and communicate effectively with diverse audiences.

Evaluation involves methodological science and craft and organizational art. Evaluation involves the selection and implementation of systematic, valid, and appropriate methodologies. Evaluation also involves the organizational establishment, management, deployment, and dissemination of a portfolio of studies and associated findings that provide defensible and relevant information to decision makers.

Since at least the 1960s, program evaluation and its close companion, policy analysis, have become institutionalized aspects of congressional oversight and agency management of federal programs.[1] Institutionalization, however, is not synonymous with acceptance, quality, credibility, or impact.

Many federal and state agencies struggle to implement evaluation programs, which may be mandated but may, in fact, be unwelcome appendages to program operations and/or organizational decision making. Evaluation efforts are some-

[1]Joseph S. Wholey, John W. Scanlon, Hugh G. Fukumoto, and Leona M. Vogt, *Federal Evaluation Policy, Analyzing the Effects of Public Programs* (Washington, DC: The Urban Institute, 1970); A. Melt2sner, Policy Analysis in the Bureaucracy (Berkeley, CA: University of California Press, 1976).

times undertaken only when required by outside forces, such as legislation. These realities make the evaluation program of one of the nation's public-private partnership programs, the Advanced Technology Program (ATP), all the more striking. The design and implementation of its evaluation program offers a creative laboratory for learning more about evaluation, particularly in the field of science and technology.

This report addresses the science, craft, and art of evaluation in the context of ATP, a program within the U.S. Department of Commerce, National Institute of Standards and Technology (NIST). In terms of science and craft, this report describes the evolving set of methodological techniques ATP has used to monitor and assess programmatic impacts, and reports methodological and empirical advances generated by ATP. The report assembles a large body of past work into a coherent framework, making it more accessible to a diverse audience.

In terms of art, the report describes the creation and evolution of ATP's evaluation program. It describes how the program has used evaluation techniques to answer questions directed at its fundamental rationale, design features, and economic impacts.

This report also highlights ways in which early challenges to the program led to an evaluation program noted for its methodological variety, recourse to nationally prominent scholars, and a distinctive emphasis on disseminating its findings in peer-reviewed literature and in policy-relevant presentations for agency officials and political constituencies. Finally, the report shows that in politically contested arenas, methodological rigor and empirically grounded findings are necessary but not sufficient to protect or advance a program. For even as ATP drew national recognition in the evaluation community for its systematic, and rigorous evaluation program, congressional critics of the program repeatedly charged that the program lacked adequate evidence of its impacts.

The report is presented in three major parts, plus this introductory section. The remainder of the introduction presents the political underpinnings of ATP and discusses further the role of evaluation. Part I provides a general framework for evaluation, and in the context of the framework, discusses evaluation fundamentals and methods, best practices, and ATP's evaluation program. Part II demonstrates the use of evaluation methods by citing, both through direct quotation and paraphrase, the contents of a selection of ATP evaluation studies. Part III presents the emerging body of knowledge from the studies of ATP over the past decade—knowledge about evaluation, firm behavior, and ATP's performance. Part III also presents conclusions and recommendations, summarizing key points and identifying remaining questions, issues, obstacles, and challenges, and promising opportunities to learn more through evaluation.

The Political Economy of the Advanced Technology Program

ATP is an outgrowth of a national policy dialogue directed at redressing systematic gaps in the market settings that link scientific advances to technological innovation. ATP was designed to fill those gaps by a project selection and funding process that fosters and enhances new and intensified modes of collaboration among private, not-for-profit, and public sector organizations engaged in high-risk research, all with the objective of accelerating the development and commercialization by U.S. firms of enabling technologies.

ATP was established in 1988 under Title V (Technology Competitiveness Act), Subtitle B of the Omnibus Trade and Competitiveness Act (P.L. 100–418), and received its first appropriations in fiscal year 1990. The program is a response both to a specific period in United States international economic competitiveness and to a longer-term historic perspective of the federal government's contribution to the development and commercialization of potentially significant technological advances of a broadly enabling nature.

The specific historic backdrop for ATP's creation was a pervasive concern in the United States throughout much of the 1970s and 1980s, documented by a variety of key economic indicators, that the nation's slow rate of economic growth and worsening international trade balance were attributable in part to its loss of technological competitiveness. The U.S. innovation system was widely seen as exhibiting structural flaws. Points of concern included a loss of international and domestic markets in technology-intensive products, a failure to gain market position for products U.S. firms had helped research and develop, and a faltering standing in R&D "races" to exploit the commercial significance of emerging scientific advances.[2]

The flaws were seen as products of an emphasis in U.S. policies on scientific leadership, mission-directed R&D, and breakthrough discoveries, paired with a lack of attention to technology development and deployment. Critics believed an overemphasis in these areas was made at the expense of diffusion-oriented strategies and programs that would assist U.S. firms and laboratories to gain technologically and economically when their scientific advances were converted into new and improved products and processes.[3] Critics argued that the United States needed an "innovation policy" rather than a "science-only policy."

These flaws were described in both absolute and relative terms. In an absolute sense, the flaws were seen as reflecting specific forms of market failure, primarily the lack of adequate incentives for private firms and private capital markets to fund "high-risk, high-return research on broadly enabling, precompetitive technologies for which appropriability was at issue."[4] In a relative sense, the challenge to the United States was that major economic competitors, most notably Japan and some members of the European community, were held to be far more willing to use national funds to support the development and commercialization of civilian technologies, particularly those seen as providing "first mover" advantages in new, strategic, or large commercial markets.[5] Lack of

[2]U.S. Congress, Office of Technology Assessment, *Making Things Better: Competing in Manufacturing,* OTA–ITE–443 (Washington, DC: U.S. Government Printing Office, 1990).

[3]H. Ergas, "The Importance of Technology Policy." In P. Dasgupta and P. Stoneman, eds., *Economic Policy and Technological Performance* (New York: Cambridge University Press, 1987), pp. 51–96; R. Florida and M. Kenney, *The Breakthrough Illusion* (New York: Basic Books, 1990).

[4]L. Branscomb and G. Parker, "Funding Civilian and Dual-Use Industrial Technology." In L. Branscomb, ed., *Empowering Technology* (Cambridge, MA: MIT Press,1993), pp. 64–102.

[5]U.S. Congress, Office of Technology Assessment, *Competing Economies: America, Europe, and the Pacific Rim,* OTA–ITE–498 (Washington, DC: U.S. Government Printing Office, 1991).

comparable government programs in the United States was held to place U.S. firms at a disadvantage, because they had to bear the full burden of supporting costly and risky R&D projects with questionable appropriability of profits.

The design of ATP also drew upon emerging perspectives about the character of competition among firms. Knowledgeable observers no longer saw firms that conducted business in the same product lines as engaged exclusively in Darwinian struggles for survival, and they no longer identified firms engaged in buyer-seller relationships only as seeking to maximize profits/minimize costs from one-time transactions. Instead, both theory and increased documentation of business practice pointed to numerous forms of collaboration between such pairs.[6]

In particular, with respect to R&D, it became increasingly evident that firms sharing a common interest in selected "generic" technologies could increase their competitiveness through collaboration. IBM's agreement in 1992 to join with Toshiba of Japan and Siemens of Germany to develop memory chips illustrates what was seen as a new way of doing business, especially in R&D-intensive and technology-driven sectors.

Collaboration spread the capital costs of large-scale R&D projects among investors. It offered a way to explore otherwise out-of-reach technological frontiers, both as part of a firm's offensive strategy of finding new products and markets, and as part of a defensive strategy of better predicting and understanding the rise of disruptive technologies that threatened a firm's corebusinesses.

The growing acceptance of inter-firm and inter-sector collaboration, especially in pre-competitive, generic (or enabling) R&D, was reflected in several legislative changes, such as the National Cooperative Research Act of 1984,[7] that relaxed antitrust bars to collaborative R&D programs. The evolving framework also saw a relaxation of the analytical and ideological categories that had previously been

[6]D. Mowery, ed., *International Collaborative Ventures in U.S. Manufacturing* (Cambridge, MA: Ballinger, 1988).

[7]A. Link and L. Bauer, *Cooperative Research in U.S. Manufacturing* (Lexington, MA: Lexington Books, 1989).

used to define the boundaries of federal government and private sector roles in supporting R&D directed at civilian-oriented technology.[8]

Under the framework that dominated policy thinking following World War II, the federal government had responsibility for funding basic research and R&D related to mission-oriented national objectives such as defense, while the private sector had responsibility for R&D directed at civilian-oriented products and processes. Reinforcing the hold of this paradigm was the checkered political and technological history of efforts by the federal government to promote specific technological innovations.[9, 10] Over time, however, the paradigm began to yield to a more complex but nuanced appreciation of both the overlap and holes in these seemingly fixed boundary markers.[11] In particular, the emerging model emphasized government support of technically risky projects that had prospects for technological and commercial importance; significant amounts of industry cost sharing; sunset requirements for government funding; selection of projects for funding in a fair and open competitive process free of political influence; large spillover effects; and collaboration with other firms and organizations.

Despite growing support for public-private partnerships emphasizing high-risk, enabling technology, opposition continued.[12] Opposition embodied several

8B. Smith, *American Science Policy Since World War II* (Washington, DC: Brookings Institution, 1990).

9See L. Cohen and R. Noll, *The Technological Pork Barrel* (Washington, DC: Brookings Institution, 1991) for examples of inefficient government funding projects and a discussion of factors behind the inefficiencies.

10See Charles W. Wessner, ed., *The Advanced Technology Program: Challenges and Opportunities* (Washington, DC: National Academy Press, 1999), pp. 1–4, for a summary description of the instrumental role played by the federal government since its earliest history in the development of new production techniques and technologies. Examples include a government contract in 1798 with Eli Whitney to help lay the foundation for the U.S. machine tool industry, government funding to demonstrate the feasibility of Samuel Morse's telegraph, and government support for the development of the radio.

11National Academy of Sciences, *The Government Role in Civilian Technology* (Washington, DC: National Academy Press, 1992).

12C. Hill, "The Advanced Technology Program: Opportunities for Advancement." In L. Branscomb and J. Keller, eds., *Investing in Innovation* (Cambridge, MA: MIT Press, 1998), pp. 153–173; P. Hallacher, "Effects of Policy Subsystem Structure on Policymaking: The Case of the Advanced Technology Program and the Manufacturing Extension Partnership Program," Ph.D. Dissertation, Pennsylvania State University, 2000.

reinforcing ideological and theoretical positions.[13] It started with the premise that the national government contributes most effectively to long-term economic growth and technological innovation when it adopts a minimalist approach to involvement in the economy, focusing its activities on the enforcement of private contracts, the provision of stable monetary aggregates, the maintenance of certain and low taxes, and the provision of a small, carefully delimited set of selected public goods.[14] Opposition was also based on propositions that public sector support cannot increase the rate of commercially productive R&D because of the phenomena of moral hazard, adverse selection, rent-seeking behaviors by firms applying for support, and bureaucratic entrepreneurship on the part of agency managers pushing for overly rapid development of untested and economically problematic technologies.[15]

Opponents of federal domestic technology development programs also point to further defects. First, they believe the programs may simply substitute public-sector R&D dollars for private-sector R&D dollars. Second, they say these programs may have untoward distributive effects, unfairly benefiting some U.S. firms and/or industries at the expense of others, such as large firms at the expense

---

[13]Opposition extended to earlier, contemporary, and subsequent similar programs that involved efforts by the federal government to stimulate acceleration of commercially oriented technological innovation through targeted selection of industries, firms, technologies, or projects. See B. Smith, *American Science Policy Since World War II,* 1990.

[14]Federal government support of basic research rests on a broad-based, bipartisan political consensus that it is a public good that would not be adequately provided by the actions of the private sector alone, and that it contributes in significant ways to national objectives in defense, health, space, and economic competitiveness.

[15]"Moral hazard" refers to the actions by individuals or firms to increase their risk taking behavior in response to the existence of insurance or other forms of compensation for costly outcomes. Thus, while ATP purposefully underwrites a portion of an R&D project's risk to encourage private firms to take on more challenging—hence riskier—technical problems, the phenomenon of moral hazard raises the question of whether or not there is a side effect of promoting inefficient resource use. "Adverse selection" refers to a problem arising from asymmetry in the quality of information possessed by the applicant firms and ATP about the riskiness of prospective R&D projects. If ATP knows less about the risks than the companies, it will be at a disadvantage in its selection decisions. "Rent-seeking behaviors" refers to efforts directed at acquiring or making permanent a stream of payments, government program, or regulatory arrangement that yields returns to an economic group above that which they would receive under competitive market conditions. Such behavior would occur if awardees attempted continuation of their government awards beyond the scheduled end date or if political efforts were taken to maintain the program even if it were deemed not to produce its intended results.

of small firms. Additionally, opponents claim the programs may respond to pressures from bureaucratic and technological interests, and unduly push for rapid deployment of technologies before the technological feasibility or market demand has been demonstrated. They maintain that these programs may fail to select proposals with the highest likelihood of achieving technical success, commercial success, and large spillovers. They also maintain that funding a project creates vested interests that causes its funding to become self-perpetuating. Finally, opponents say the claims about the productivity of collaborative efforts, and the federal government's role in fostering them may be challenged as unproven.

The catalog of arguments on behalf of the benefits of public-private partnerships in general, and ATP in particular as well as the catalog of possible flaws are both mixes of normative perspectives on the role of the public sector, and theoretically and empirically testable propositions. The competing catalogs, in effect, describe domains of debate and decision making in which evaluation may contribute to improved public policy.

The Role of Evaluation

Evaluation has a recognized and well-understood role in the operations of most public sector agencies. As phrased by Mark, Henry, and Julnes:[16]

> Evaluation assists sense making about policies and programs through the conduct of systematic inquiry that describes and explains the policies' and programs' operations, effects, justifications, and social implications. The ultimate goal of evaluation is social betterment, to which evaluation can contribute by assisting democratic institutions to better select, oversee, improve, and make sense of social programs and policies. (p. 3)

Evaluation has served all these purposes for ATP: to oversee, improve, and make sense of the program—both within and outside of NIST. ATP's evaluations have helped program administrators revise and refine the program so that the program is harmonized with legislative intent. Evaluation has been used to respond to

16M. Mark, G. Henry, and G. Julnes, *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs* (San Francisco: Jossey-Bass, 2000).

Congressional and OMB questions about program characteristics and impacts. ATP's evaluation program has had a broader span of coverage than typically encountered across federal agencies. Its studies have ranged from monitoring and surveying the characteristics of early grantees to sophisticated theoretical and econometric efforts designed to tease out differences between the performance of ATP awardees and non-awardees. In pursuit of these multiple objectives, ATP has strategically employed an evolving and broad set of methodologies.

Early ATP evaluation activities focused on measuring progress, generating information about awardees, and projecting economic impact. In the context of political opposition to ATP, the program's evaluation agenda has often been shaped in response to ongoing challenges, such as whether ATP awards were being used by firms to displace private sector funds. Over time, as some funded projects have had time to progress through technical research and commercialization stages, ATP's evaluation program has centered more on measuring impacts. These include both private impacts garnered directly by the firms participating in ATP projects, and spillover impacts that are one of the justifications for the program.

Finally, evaluation has served as a test of two of ATP's core and linked premises; that is, that a federal agency program can strategically and selectively generate additional and innovative modes of R&D collaboration between and among organizations, and that this collaboration among organizations will accelerate the rate of technological innovation in the nation's economy. ATP's evaluations have provided new insights into the characteristics of workable collaborations among various sectors, the character of gaps between scientific discovery and product development, the strategies and behaviors of firms as they seek to close or bridge these gaps, and the forms and networks through which knowledge spillovers diffuse within an economy.

PART I:

EVALUATION FRAMEWORK

Developing a solid evaluation program starts with a logical framework. The framework links the evaluation activities to the program it serves. This logical framework guides the construction over time of a useful portfolio of evaluation studies informative to stakeholders. The purpose of Part I of the report is to discuss and illustrate the elements of an evaluation framework. It will serve to anchor the remainder of the report.

Part I is presented in three chapters. Chapter 1 presents a general framework showing how evaluation relates to a program, its mission, and its intended results. It also provides an overview of selected evaluation principles and best practices. Chapter 2 reviews methods of evaluation and explains their special uses, advantages, and disadvantages. Chapter 3 moves from the general to the specific, presenting a logical framework, or logic model, of ATP's evaluation program.

For the general reader, the ATP framework illustrates how an effective evaluation program is developed in tandem with the program it serves. For the ATP staff, the program's evaluation framework provides a context for reviewing past studies it has commissioned, and for planning future studies.

CHAPTER 1

Evaluation Fundamentals

Objectives of Evaluation

Evaluation is the systematic investigation of the value or merit of a thing or an activity. Evaluation has a long history, reportedly dating back 4,000 years to China, where it was used to assess public programs. While evaluation often is viewed as an adversarial process, it can also be viewed as a tool that not only measures, but also contributes to success.

Program evaluation looks at the impacts of a collection of projects. Project evaluation focuses on individual projects. Project evaluation can start sooner than program evaluation, which requires time for multiple projects to make progress toward program goals.

Evaluation can tell an organization what is working and what is not. It can reveal what outputs are being produced, and with what efficiency. It can indicate when performance is improving and when performance is declining. It can answer a host of questions needed to determine if an activity or suite of activities is on track to produce desired outcomes. Evaluation can also address "why" questions. In sum, evaluation can be as much a method of learning as it is a method of documentation, and certainly far more than a mandated obligation.

Table 1–1 summarizes major reasons for evaluating programs and activities and gives examples of the kinds of information that may be developed in pursuit of each purpose.

Table 1–1. Why Evaluate?

PURPOSE	EXAMPLE
For internal program management	Identify program activities that address key program goals and outcomes and those activities that do not
To answer stakeholder questions	Determine if the intended beneficiaries receive net benefits
To meet official requirements	Report program metrics on inputs, outputs, and outcomes as required for GPRA* reporting
To understand specific phenomena	Assess factors that determine effective collaborations
To promote interest in and support of a program or activity	Make study results available through journal articles, reports, press releases, and presentations

*GPRA is the abbreviation for the Government Performance and Results Act of 1993, which is further discussed below.

For Internal Program Management

Managers who are not continually assessing their operations lack the information to manage strategically. Systematic monitoring, analysis, and feedback of information enable managers to identify and implement needed changes.

To Answer Stakeholder Questions

All organizations are accountable to someone, and the persons or organizations to which they are accountable are called stakeholders. Stakeholders generally offer their support to a program on the condition that their targeted goals and objectives are met, and stakeholders usually require evidence that they are getting the desired return on their investment. Stakeholders in private organizations are stockholders and other owners and investors, with employees, managers, and boards of directors also in stakeholder roles. While stakeholders in public agencies are ultimately the citizens, the legislative and executive branches are the *de facto* stakeholders. Other individuals and groups, including agency employees, program administrators, oversight committees, program participants, and the communities from which these participants are drawn, also have stakeholder roles. Evaluation can be used as a tool for stakeholders to learn if their objectives are being met.[17]

To Meet Official Requirements

Steps are sometimes taken on stakeholders' behalf to institute specific evaluation and reporting requirements. In 1994, for example, the U.S. Congress passed the Government Performance and Results Act of 1993 (GPRA), requiring, among other things, all federal agencies to develop strategic plans that identified agency objectives, to relate budgetary requests to specific outcome goals, to measure performance, and to report on the degree to which goals are met. The requirements require administrators to move beyond a focus on activities to a focus on results. The GPRA is aimed at improving efficiency, effectiveness, and accountability in federal management and budgeting.[18]

In a study of the feasibility of implementing the GPRA for basic and applied research, the Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, took the following position with respect to applied research:[19]

> ...reporting on applied research...consists of systematically applying methods widely used in industry and in some parts of government. For example, an applied research program usually includes a series of milestones that should be achieved by particular times and a description of the intended final outcomes and their significance. Periodic reporting can indicate progress toward those milestones. (p.2)

GPRA now drives the evaluation efforts of many federal agencies. Some state government legislations have also put into place requirements for evaluation of

---

[17]Audits and financial reporting requirements are other mainstay tools used to protect stakeholder financial interests, but they are not the focus of this report.

[18]An overview of the GPRA is provided in Appendix 1 of the General Accounting Office Executive Guide, "Effectively Implementing the Government Performance and Results Act," GAO Report GGD–96–118, Washington, DC, 1996.

[19]National Academy of Sciences, Committee on Science, Engineering, and Public Policy, *Evaluating Federal Research Programs; Research and the Government Performance and Results Act* (Washington, DC: National Academy Press, 1999).

state programs.[20] As a result, more public programs are subjected to evaluation, and more data on their operations and results are becoming available.

In 2002, the Bush Administration added additional requirements related to program evaluation aimed at improving performance management practices of federal agencies.[21] The President's Management Agenda includes new investment criteria and a program assessment rating tool, known as PART, for federal R&D programs

To Understand Specific Phenomena

A program frequently encompasses a complex array of theories, issues, and perspectives. Evaluation of the program thus may present an opportunity to investigate a wide range of phenomena of interest. For example, ATP's legislative charge to foster collaborative activities provides a singular opportunity to explore various aspects of collaboration. Evaluation thus helps ATP benefit from more knowledge about the strengths and weaknesses of collaborations as well as others who use collaboration to accomplish goals.

To Promote Interest in and Support of a Program or Activity

Evaluations inform and educate a large audience. Most audiences wish to hear not just about goals and objectives but also about accomplishments. Administrators of both public and private programs who must make and defend budget requests and solicit support are increasingly expected to present well-documented evidence that past budgets are producing desired results and that there are reasons to expect that future budgets will also pay off. Public policy makers look for evidence that programs are, or are not, working.[22] Over the long run, public opinion and underlying support can be influenced by evaluation-based evidence

[20]See, for example, Susan Cozzens and Julia Melkers, "Use and Usefulness of Performance Measurement in State Science and Technology Programs," *Policy Studies Journal,* 25: 425–435, 1975.

[21]Office of Management and Budget, *The President's Management Agenda,* available online at http://www.whitehouse.gov/omb.

[22]Politics and ideology also play major roles in determining support for a particular effort. Here the focus is on the objective basis of support.

that specific programs work or do not work. In contested political environments, pro and con evaluations may compete for public attention. The quality and credibility of the evaluations thus can become part of the larger competition for public acceptance and support. Program administrators need to be able to explain their core competencies and their unique contributions and to document their claims of impact with the best possible evidence.

Mapping Evaluation to Mission and Stages of Implementation: A Generic Evaluation Logic Model

A recommended starting point in planning evaluation is to develop a logic model.[23] A logic model is intended to provide a clear diagram of the basic elements of a program, subprogram, or project, revealing what it is to do, how it is to do it, and with what intended consequences. It shows the logical linkages among mission, activities, resources (inputs), outputs, outcomes, and impacts. It is a first step in identifying critical measures of performance. The logic-model tool has been used in program evaluation more than 20 years ago and has been adapted to program planning where it helps ensure a correspondence among all the elements of a program.[24]

Evaluation works best when it is closely mapped to a program's mission and to each stage of program implementation through a logic model. Figure 1–1 shows a generic evaluation model that reveals the integral relationships of a hypothetical public program and its evaluation program. This evaluation logic model is tailored to assess effects at each stage and to provide feedback from evaluation to program administrators and policymakers.

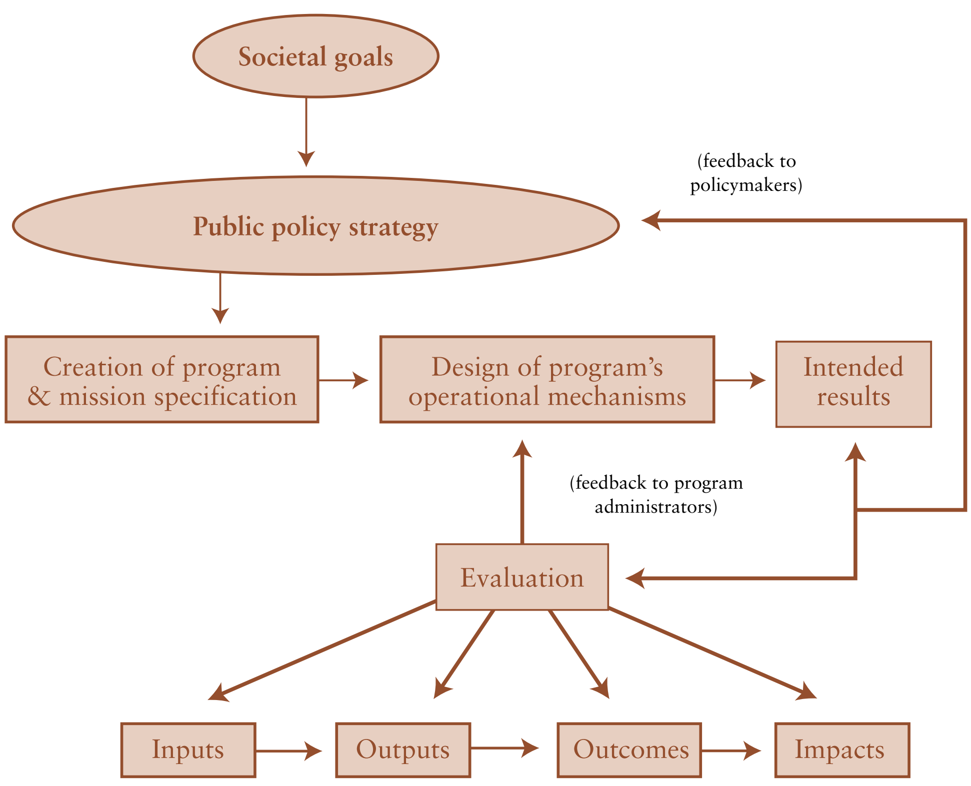
---

[23] A number of guides to logic modeling are available. For a listing of guides that focus on the forms and structures, strengths and weaknesses, and uses of logic modeling, see Molly den Heyer, ed., *A Bibliography for Program Logic Models/Logframe Analysis* (Ottawa, Canada: Evaluation Unit, International Development Research Centre, 2001). Included in the list is John A. McLaughlin and Gretchen B. Jordan, "Logic Models: A Tool for Telling your Program Performance Story," *Evaluation and Program Planning,* 22:65–72, 1999.

[24] Paul F. McCawley, *The Logic Model for Program Planning and Evaluation,* CIS 1097 (Moscow, Idaho: University of Idaho Extension Program), p. 1.

Starting at the top left of Figure 1–1 and working down and across, a societal goal, such as economic prosperity or improved health, provides the impetus for establishing a public program. The program is then presumably designed to carry out the intended mission, and it must obtain resources to carry out its mission. These resources, or "inputs," expressed in monetary terms, convey the program's costs to the public. A program's structure and operational mechanisms for carrying out its mission determine how the inputs are used and what they produce in terms of program "outputs." For example, a program driven by the societal goal of improved health might fund research in infectious diseases. Program appropriations allow it to purchase labor and materials needed for oper-

Figure 1–1. Evaluation Logic Model



Note: The dynamics by which the transformation among these various stages occur are often complex and are themselves the subject of evaluation.

ations. Short-run program outputs might include publications, presentations, workshops, and test results. Next-stage outputs might include prototype therapies, and prototype vaccines, followed by clinical trials. Longer-term program outcomes might include treatments and vaccines applied by medical establishments. Long-run impacts might include reduced rates of disease spread, higher survival rates of those infected, and reduced mortality rates for the nation. All of the impacts of a program should be assessed against the program's mission, intended results, and costs. A final, important step is to feed the evaluative findings back to inform program administrators and policymakers and to improve the program's structure and operations.

Evaluation also looks at the process dynamics whereby inputs are converted to outputs, which in turn may lead to outcomes, which in turn may translate into impacts. Evaluations made at various stages can determine how a program is progressing toward its mission. To continue the above health example, evaluative techniques may be used to assess the transfer of the program's research results to pharmaceutical companies. Evaluators might explore how the rate of transfer can be accelerated. They might ask what percentage of medical establishments is using program-derived improved treatments and vaccines at a given time. They might investigate how the program affects production costs, what the likely outcome would have been without the program, and what return society is likely to receive from investing in the research program. From these investigations, insights may be gained for modifying the program in ways to bring its dynamics in closer accord with its mission.

Evaluation can provide answers to these and other questions that arise during a program's implementation and operations. Figure 1–1 suggests the complexity and scope of a full-fledged, fully integrated evaluation program.

Evaluation Steps

Regardless of their degree of comprehensiveness, most evaluation studies can follow a systematic procedure. Table 1–2 summarizes the steps that are normally involved in organizing and carrying out an evaluation study. In practice, a study advances in an iterative manner, as previous steps tend to be revisited as a study proceeds.

Table 1–2. Steps in Evaluation

- ✓ Identify purpose and intended audiences
- ✓ Formulate questions and hypotheses
- ✓ Determine available resources, time requirements, and appropriate level of effort
- ✓ Choose evaluation method(s) and analyst(s)
- ✓ Design information gathering approaches and select or develop specific models
- ✓ Compile information
- ✓ Exercise evaluation method(s) using collected information
- ✓ Analyze and interpret results
- ✓ Write report
- ✓ Disseminate results: present to intended audiences and provide feedback to the program

Best Practices

Study Scope and Rigor

Evaluation appropriately comes in many forms. It can be comprehensive, encompassing the design, conceptualization, implementation, and impacts of a program. It can be tailored to selected features of a program. It can be focused on measurement of specific outputs or outcomes, or constructed broadly to permit tests of competing causal linkages. It can be "rigorous" in the sense of searching for the most comprehensive and systematic set of causal linkages between and among variables, employing carefully constructed and sifted data. Or it can be "good enough," that is, offering a defensible answer sufficient to the question at hand, given often severe constraints on time, budget, and access to data.[25]

[25]P. Rossi and H. Freeman, *Evaluation: A Systematic Approach,* 4th ed. (Beverly Hills, CA: Sage, 1989); Shadish et al., *Foundations of Program Evaluation*, 1991.

Designing Appropriate Tests of Program Success

It is important to define what constitutes program success before designing evaluation metrics, success being the degree to which a program produces its desired outcomes. That a program's founding legislation often contains multiple desired outcomes, some of which may be contradictory or entail sizeable tradeoffs, is a common conundrum. At the least, multiple tests of success may be needed, as is the case with ATP.

How to measure success is itself one of the decisions surrounding the design of an evaluation program. Rossi and Freeman, for example, in their treatment of measuring efficiency, list 15 key concepts.[26] Included in the list are standard measures such as benefits, cost effectiveness, cost-benefit ratio, distributional effects, externalities, opportunity costs, and shadow prices. Specification of the measure(s) of success can influence the selection of evaluation design, identification of data to be collected, type of statistical or other tests of causation or significance to be employed, and, ultimately, conclusions about program impacts.[27] (See page 67 for discussion of specific tests used to assess ATP's success.)

Use of Control Groups and Counterfactuals

Evaluation seeks not only to measure change, but also to determine if the cause of change is attributable to program intervention. Evaluation thus is directed at ruling out alternative, competing explanations for the change.

Indeed, alternative explanations for observed changes frequently abound. To cite a simple but important example, a program begun at a cyclical trough may be

[26]P. Rossi and H. Freeman, *Evaluation: A Systematic Approach,* p. 377.

[27]Though it may seem self evident, stakeholders sometimes need to be reminded that a program, to be fairly tested, must be measured against its mission. This is sometimes overlooked, inadvertently or deliberately, either to the detriment or benefit of the program in question. For example, ATP was once criticized that it had "merely accelerated technology development," when, in fact, accelerating technology development is an accomplishment that is core to ATP's legislated mission. This criticism of the program's accomplishments is therefore invalid. Being able to link the measured effects directly back to mission is critical for program administrators who present and defend a program. Similarly, when presenting unintended results, it is incumbent on program administrators to acknowledge that the results presented, though they may be desirable, are not within the mission scope.

followed by economic improvement on the part of a firm or region that participates in the program, but this improvement could plausibly have been associated with the economy's general recovery. A well-designed and implemented evaluation, however, may show that the program caused the positive impacts. This ability to "isolate" or "demonstrate" cause is one of the most important tasks evaluation performs for an agency.

To rule out alternative explanations, it is often necessary to contrast the changes that occurred in the group participating in the program with a comparable or "like" group. Comparison and control groups are generic techniques used to gauge whether the observed changes would have occurred even without the program. In each case, evaluators seek to find a population held to be like the participating population in all relevant respects other than participation.[28]

Random assignment of participants to either experimental or control conditions is an approach that allows evaluators to generate two groups that have the same general characteristics in all salient variables other than the one to be tested. This is the approach generally used, for example, in testing the effectiveness of medical treatments.

Whether projects in a public program are selected randomly, by merit, or by other criteria, evaluation requires the construction of a comparable group. The critical decision is the selection of what Mohr has termed the "criterion" population.[29] To whom should program participants be compared? Non-funded applicants? All firms in the same technology/industry sector? All firms? As Judd and Kenny note, in referring to different ways in which the population may be tiered, "Deciding on which level is the most appropriate in any piece of applied research is a fundamental and frequently difficult problem."[30]

[28]D. Cook and J. Stanley, *Experimental and Quasi-Experimental Design for Research* (Chicago: Rand McNally, 1966).

[29]L. Mohr, *Impact Analysis for Program Evaluation,* 2nd ed. (Thousand Oaks, CA: Sage, 1995).

[30]Charles M. Judd and David A. Kenny, *Estimating the Effects of Social Interventions* (New York: Cambridge University Press, 1981), p. 55.

After an evaluator chooses the criterion population, the next decision becomes how members of this population are chosen for the comparison group. Random assignment can work if the criterion population is large enough, but other considerations may lead to a more purposeful selection of a "matched" set of actors. For instance, say an agency can fund only one half of its eligible applicants, and awardees and non-awardees are alike in program-relevant criteria. In this case, the impacts of the program will be better measured by comparing the technical and economic performance of these two groups rather than comparing the program's participants to a random selection of all firms within the relevant technology/industrial sectors.[31]

To take things further, an evaluator might assess the program's impacts by comparing the awardee/non-awardee performance as described above, and also comparing each of those groups to a randomly selected set of firms from the criterion population. Comprehensiveness of design, however, comes with a price. The larger and more differentiated the comparison and/or control groups, the more expensive the project and the more complex the conduct of the evaluation.

An alternative approach, termed a "counterfactual," solicits expert judgment about what would likely have happened in absence of the program in question. For example, had a program not funded the development of a new technology, would someone else have developed it? In the same timeframe? With the same features? By the same parties? If it had not been developed, what would have been used in lieu of the new technology? Or if it had been developed differently without the program, what would the differences likely have been? Expert judgment about the most likely alternative scenario is used to establish a base line against which the program result can be compared. Generally the use of a counterfactual to isolate the effect attributable to a program is less rigorous than use

[31] Randomization is less useful—indeed it is likely to be politically unacceptable if used for project selection—when an agency seeks the "best" of a set of proposals. Phrased differently, it would not be expected that an agency announce a program competition, set criteria, and then randomly select grantees from among a set of applicants. That agency would instead measure the applicants against the specified selection criteria. An "intermediate" process does exist: An agency could announce a program competition, convene selection panels to sort proposals into acceptable/non-acceptable categories, and randomly select awardees from the acceptable category. No federal agency has yet been willing to experiment with the technique to the authors' knowledge.

of a control or comparison group. But, it has the advantages of often being feasible when a control or comparison group is not, and it is a respected approach in social sciences research, where controlled experimentation is usually difficult, if not impossible.

Deciding Who Should Perform Evaluation

The skills to perform evaluation are diffused throughout several sectors of the economy. As attested to by the large and diverse membership of the American Evaluation Association and the existence of several evaluation journals, evaluators come from the fields of economics, sociology, history, statistics, public administration, and other fields. The nature of the study and the methods to be used influence the selection.

The use of internal or external evaluators is often an important decision; each choice has advantages and disadvantages. Evaluation by in-house staff offers the following several advantages: staff evaluators can make important contributions to mission-focused studies; provide continuous feedback between evaluation staff, operations staff, and agency decision makers; and translate findings into agency-relevant publications and briefings. Using staff evaluators facilitates access to and use of confidential data related to private sector activities that are controlled by federal agencies. Perhaps more important than any of the above, command of the analytical and empirical skills to conduct evaluation is essential to an agency's ability to design, manage, monitor, interpret, assess, and disseminate studies and findings performed by outside contractors.

Sole reliance on in-house staff, though, runs the risk of neglecting newer conceptual, methodological, or empirical advances occurring in program evaluation in other organizations. It also runs the risk of inconsistent, incomplete evaluations when agency personnel are detailed for other pressing agency needs. Legislatively imposed constraints on personnel classifications, staffing limits, and salaries may simply make it impossible for an agency to assemble the staff required to conduct the range and depth of an evaluation program deemed necessary. But, most importantly, relying totally on an in-house staff for evaluation runs the risk of reducing credibility for the studies. In short, especially in politically contested domains, recourse to external evaluators is as much a matter of political credibility as of comparative methodological expertise.

Use of outside contractors can be an indispensable component of an agency's evaluation efforts, Reliance, however, on outside evaluators presents its own set of problems. External evaluators may not know the ins and outs of a program and may not want to invest the time to learn. External evaluators may have their own agendas, which may place more importance on pursuing a particular line of research than on addressing questions of interest to program administrators. External evaluators may see their audience in very different terms than do program administrators, and may resist translating their studies to language that program stakeholders can understand. Program participants who are reluctant to share data with "outsiders" may meet external evaluators with distrust. Finally, external evaluators may be lacking in objectivity and credibility if, for example, they have established their reputations based on particular findings, such as that all federal programs are either wasteful or efficient, or that a particular technology or industry is of prime importance or of little importance.

Transparency and Replicability

Good evaluations follow the dictates of good scientific research; that is, good evaluations set forth explicit hypotheses and research protocols, make public the documentation used, and delineate the tests of impact or causation, statistical or otherwise, used to formulate conclusions. Holding aside possible differences in interpretation that may follow from two or more researchers examining the same data and tests, a key to a sound evaluation is that its inner workings are transparent, thus permitting others to replicate it if they choose, or to adjust the evaluation's mechanics to determine how sensitive the findings and recommendations are to specific assumptions.

Other Considerations

Codes and criteria for good evaluation practice abound, albeit flexibility rather than orthodoxy characterizes the best and most recent outlooks of nationally prominent evaluators.[32] At a minimum, best practice means: (1) addressing significant programmatic questions, (2) linking evaluation questions and design to

[32]Mark et al. *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs,* 2000.

program mission, (3) focusing on outputs and outcomes rather than only inputs, (4) carefully identifying and collecting relevant data, (5) considering alternative explanations for observed changes, (6) ensuring transparency in treatment of assumptions and presentation of data and other evidence, (7) using a degree of methodological rigor and care that can withstand critical scrutiny, and (8) communicating the findings effectively.

Evaluations are often research-oriented undertakings, designed to test hypotheses and generate primary data. Such efforts need to be conducted with the same level of precision that would follow in seeking publication in high-quality, peer-reviewed journals. Program administrators anxious for bottom-line results may meet the research nature of some evaluation studies with impatience. Evaluation staff will need to recognize the possible tension between pursuing research-oriented, exploratory studies, which advance the tools of their trade, and producing and communicating studies of immediate applicability. Finding a workable balance is important to achieving both short- and long-run success for an evaluation program.

CHAPTER 2

Choosing Methods of Evaluation

Evaluators use a variety of methods to address questions of program performance. The methods share common features, but each has its advantages, disadvantages, and specialized purposes. Former methodological wars about the relative merits of different techniques have largely given way to an eclectic approach in which techniques are chosen for their appropriateness to the evaluation question at hand, to cost and administrative feasibility, and to a purposeful mixture of methodological paradigms.[33]

The use of some methods of evaluation depends on how a program is positioned relative to the market and how mature it is. Some methods are particularly useful in assessing early-stage research programs, while others are better suited for assessing later-stage, closer-to-market programs. Both the utility and feasibility of methods may change as a program develops. Generally, the more an R&D program's scope spans from research to commercialization, the more methods evaluators can use to capture the full range of a program's impacts. Recognizing this, a farsighted strategy would be to design an evaluation program that lets data perform multiple duties. For instance, early evaluations may be designed to generate survey information on participants for immediate use, with the idea that the survey information can later be used as baseline information for subsequent evaluations. A farsighted strategy would also use multiple methods to capture the full range of a program's impacts, to triangulate findings on salient program impacts, and to identify and validate relationships and impacts not readily

[33]See J. Green and V. Caracelli, "Defining and Describing the Paradigm Issue in Mixed-Method Evaluation," in J. Green and V. Caracelli, eds., *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms* (San Francisco: Jossey-Bass, 1997), pp. 5–17.

apparent in the construction of the initial design. Table 2–1 lists major evaluation methods, defines each, and illustrates how each may be used to gather information about a program's performance.

Crosscutting the methods are various approaches to collecting data. Among the approaches to collecting data for use in evaluation are systematic and anecdotal observation, review of records and searches of existing databases, testing, experimenting, recording responses of focus groups, interviewing experts, and conducting surveys by structured interviews in person or by telephone, or by mailed or electronically administered questionnaires.

The remainder of this chapter provides overviews of the methods listed in Table 2–1. In-depth coverage is beyond the scope of this toolkit. Volumes could and have been written on the various methods. Readers seeking additional information are directed to references provided.

Analytical/Conceptual Methods for Modeling and Informing Underlying Program Theory

Clarifying and validating a program's underlying concepts and theories, and investigating analytical linkages among program elements, are frequently important parts of an agency's overall program evaluation strategy. Modeling and informing the underlying theory of the program is an ongoing rather than a one-time task in a dynamic program. It is an essential prelude to program operations, an early stage complement to formative management and evaluation efforts, and a building block in the design and revision of a long-term comprehensive evaluation program as experience and evidence accumulates about program impacts.

Why is this form of self-study so effective? Federal government R&D programs are typically based on a combination of hypothesized and documented relationships that link activities to objectives. Some of these relationships are explicit, evident in the language of the legislative, budgetary, and administrative debates that give rise to the program and its operational provisions. Others are implicit, involving widely shared acceptance and assumptions about stylized facts or causal linkages. These explicit and implicit relationships constitute the program's theory. As Mohr writes about program theory:

> [It] tells what is to be done in the program and why-what is to result from the program and how. It is, in short, a testable assertion that certain program activities and sub-objectives will bring about specified results. (p. 18)

In practice, however, programs are often established on partially formed theories, incomplete documentation, or fragile empirical grounds. They often reflect new, untried approaches to problems. Legislation may define program objectives without suggesting the most effective ways to meet those objectives. Furthermore, programs are frequently established on broad associations between program mechanisms and intended outcomes, implying direct, linear relationships. But the actual causal paths may be more complex; they may entail one or more intermediate steps.[34]

In addition to ideological or partisan opposition to its ends or means, a program may be challenged on the grounds that its theoretical underpinnings have logical inconsistencies, lack adequate empirical testing, and will be ineffective or inefficient in achieving intended objectives. Moreover, even where such "technical" challenges do not exist, good management practice requires that agency administrators and program officials comprehend nuanced relationships among program design features, program implementation, and outcomes.

ATP's use of analytical and conceptual methods to model underlying program theory is presented in Chapter 4.

Survey Method

Surveys can be used to describe a program in terms of frequencies, percentages, means, medians, standard deviations, and significance of sample data. Survey results are typically presented in aggregate, without identifying individual results, using tabular and graphical summaries of data. Surveys provide a statistical overview for multiple projects and participants, rather than project details, and are particularly useful in portfolio analysis.

[34]For example, Donaldson's evaluation of social programs found multiple and indirect paths that a program's actions may cause in producing intended and unintended outcomes. Stewart Donaldson, "The Theory-Driven View of Program Evaluation." Paper presented at Evaluating Social Programs and Problems: Visions for the New Millennium, February 24, 2001.

Table 2–1. Overview of Evaluation Methods

METHOD	BRIEF DESCRIPTION	EXAMPLE OF USE
Analytical/ conceptual modeling of underlying theory	Investigating underlying concepts and developing models to advance understanding of some aspect of a program, project, or phenomenon.	To describe conceptually the paths through which spillover effects may occur.
Survey	Asking multiple parties a uniform set of questions about activities, plans, relationships, accomplishments, value, or other topics, which can be statistically analyzed.	To find out how many companies have licensed their newly developed technology to others.
Case study — descriptive	Investigating in-depth a program or project, a technology, or a facility, describing and explaining how and why developments of interest have occurred.	To recount how a particular joint venture was formed, how its participants shared research tasks, and why the collaboration was successful or unsuccessful.
Case study — economic estimation	Adding to a descriptive case study quantification of economic effects, such as through benefit-cost analysis.	To estimate whether, and by how much, benefits of a project exceed its costs.
Econometric and statistical analysis	Using tools of statistics, mathematical economics, and econometrics to analyze functional relationships between economic and social phenomena and to forecast economic effects.	To determine how public funding affects private funding of research.
Sociometric and social network analysis	Identifying and studying the structure of relationships by direct observation, survey, and statistical analysis of secondary databases to increase understanding of social/organizational behavior and related economic outcomes.	To learn how projects can be structured to increase the diffusion of resulting knowledge.

Table 2–1. (Cont'd)

METHOD	BRIEF DESCRIPTION	EXAMPLE OF USE
Bibliometrics — counts	Tracking the quantity of research outputs.	To find how many publications per research dollar a program generated.
Bibliometrics — citations	Assessing the frequency with which others cite publications or patents and noting who is doing the citing.	To learn the extent and pattern of dissemination of a project's publications and patents.
Bibliometrics — content analysis	Extracting content information from text using techniques such as co-word analysis, database tomography, and textual data mining, supplemented by visualization techniques.	To identify a project's contribution, and the timing of that contribution, to the evolution of a technology.
Historical tracing	Tracing forward from research to a future outcome or backward from an outcome to precursor contributing developments.	To identify apparent linkages between a public research project and something of significance that happens later.
Expert judgment	Using informed judgments to make assessments.	To hypothesize the most likely first use of a new technology.

Conducting the Survey

Survey data can be collected by interviews conducted in person or by phone, or by questionnaires mailed, dropped off, or posted on the Internet. Questions may be either open-ended (*Why? What?*) or close-ended (*Yes/No*; or, *Which of the following choices best describes...?*). Close-ended questions may use ranking systems (*Indicate the order of your preference...*) and scales (Rate on a scale of 1 to 5).

Computation of survey statistics requires consistency across individuals in a survey group in terms of the questions asked or ranking systems or scales used. It also requires that responses to open-ended questions be coded systematically and consistently. Generally, questionnaires use a series of precisely worded, close-ended questions, and interviews use more open-ended questions and discussion, leading to more varied data that may be more difficult to analyze. But a questionnaire can also include open-ended questions, and an interview may rigidly follow a scripted questionnaire format.

Statistical inference, the process of using sample data to make inferences about the parameters of a population, reduces the time and cost of collecting data by survey from an entire population.[35] Sample design should be sufficiently described to enable calculation of sampling errors. Establishing a sampling frame—the list from which a sample is drawn—is essential. Samples may be randomized or stratified. They may be longitudinal, drawing data from the same panel of individuals at different times with the same survey questions. Or, they may be cross-sectional, drawing new samples for successive data collection.[36]

Once it has been decided that a survey method is appropriate for the evaluation task at hand, there are a number of steps involved in carrying it out. Table 2–2 lists the steps.

Advantages, Disadvantages, and Special Uses of Surveys

An advantage of using survey-based descriptive statistics in evaluation is that it provides an economical way to gather aggregate level information about a program and its participants, even in its early stages, and it accommodates the use of control and comparison groups or the collection of counterfactual information. Other advantages are that diverse audiences can usually understand the approach and results, and many people find statistical results credible and informative. Furthermore, once collected, survey data can be analyzed and reanalyzed in different ways. Surveys can provide information about participants and users not available through other sources.

[35]"Population" refers to a given, finite collection of units; "sample," to a subset of the population; and "parameter," to population characteristics. Statistics from a sample can be used to estimate unknown values of parameters for the population.

[36]There is a large literature on sampling design, which should be consulted. For additional information see, for example, Floyd J. Fowler, Jr., *Survey Research Methods* (Newbury Park, CA: Sage, 1993).

Table 2–2. Steps in Survey-Based Statistical Studies

- ✓ Lay out the objectives in detail
- ✓ Determine the required accuracy level (i.e., the acceptable sampling error)
- ✓ Identify the target respondents
- ✓ Determine the sample design
- ✓ Decide a method of collecting the data
- ✓ Design the questionnaire or interview protocol
- ✓ Conduct a pre-test to find out if everything works as intended
- ✓ Revise the questionnaire or interview guide and sampling plan, if needed
- ✓ Develop procedures for controlling response errors
- ✓ Develop procedures for follow-up with non-respondents
- ✓ Carry out necessary clerical operations to schedule and track data collection
- ✓ Administer the questionnaire or interviews
- ✓ Tabulate and analyze the data, including measures of variability, response rates, and descriptive statistics
- ✓ Write the report

There are advantages and disadvantages associated with the various methods of collecting data for descriptive statistical analysis. Phone interview works best when timeliness is important and the length of the survey is limited. Face-to-face interviews cost more and take more time, but are better for collecting complex information, using open-ended questions, providing needed flexibility, and obtaining higher response rates. Mailed questionnaires have the advantages of usually being cheaper to administer, allowing more time for respondents to form responses. The disadvantages of mailed questionnaires include relatively low response rates, a lack of flexibility, and complete reliance on the written questionnaire as self-explanatory. Web-based surveys and e-mail interactions for follow-up offer a promising approach. Using a mix-mode approach may offer advantages.

A disadvantage of survey statistics is that they do not convey the richness of individual project detail that stakeholders tend to find interesting and memorable. A further limitation is that the responses on which descriptive statistics are based are often subjective in nature. Respondents may not be truthful. They may have faulty recall. Or, they may wish to promote a particular point of view. Hence, results may be biased.

See suggested references on the survey method at the end of this chapter and examples of survey studies from ATP's experience in Chapter 5.

Case Study: Descriptive

Descriptive case studies are in-depth investigations into a program, project, facility, or phenomenon, usually to examine what happened, to describe the context in which it happened, to explore how and why, and to consider what would have happened otherwise.[37] Case studies are particularly helpful in understanding general propositions,[38] and in identifying key relationships and variables. Thus, case study can be particularly useful in the exploratory phases of a program.

Broadly Accessible Results

Most descriptive case studies are written in the narrative and are aimed at a wide audience. For example, such studies can make complex scientific and technology projects accessible to a non-scientist audience. The potential scope of the descriptive case study method is broad—ranging from brief descriptive summaries to long complex treatments.

Descriptive case studies usually start with qualitative information from direct observation, program/project documents, and interviews with key project managers. Program and project documents are useful for establishing key dates, budgets, initial plans and goals, specific outputs, key staff, and other critical information helpful in framing a study. To extend the available information, the evaluator may bring in results from one or more of the other evaluation

[37]R. Yin, *Case Study Research,* 2nd ed. (Thousand Oaks, CA: Sage, 1994).

[38]M. Shadish, T. Cook, and L. Leviton, *Foundations of Program Evaluation* (Newbury Park, CA: Sage, 1991), pp. 286–301.

methods listed in Table 2–1, such as survey results or bibliometric results, to enhance the story.

Using a "story-telling" approach, the evaluator may present the genesis of ideas, give an account of the human side of the project, explain goals, explore project dynamics, and present outcomes. Case studies can also be used to construct theories about program or project dynamics.[39] Multiple case studies may be conducted with uniform compilation of information across cases to provide aggregate statistics for a portfolio of projects.

Advantages and Disadvantages of Descriptive Case Study

An advantage of the descriptive case study method is that many decision makers read and process anecdotal cases more easily than they do quantitative studies. Another advantage is that by bringing in substantial program/project information on a less restrictive basis than most methods, case studies document events and provide a richness of detail that may prove useful in formulating theories and hypotheses that lay the groundwork for further evaluation. Case studies are also valuable in identifying exemplary or best-practice experiences. They can be used to describe how and why a program is or is not working. Used in formative evaluations, they can guide agency behavior and serve as a benchmark for other program recipients.

A disadvantage of the descriptive case study method is that the anecdotal evidence provided is generally considered less persuasive than quantitative evidence. The results of one or more individual cases may not apply to other cases.

See suggested references on the descriptive case-study method at the end of this chapter and examples of ATP's use of the method in Chapter 6.

Case Study: Economic Estimation

Economic case studies combine descriptive case histories with quantification of benefits and costs, including treatment of the distribution of benefits and

[39] K.M. Eisenhardt, "Building Theories from Case Study Research," *Academy of Management Review,* 14(4): 532–550, 1989.

costs.[40] Carrying out the descriptive analysis in advance of quantification is generally an essential step toward economic quantification. Indeed, developing an in-depth understanding of the problem in its "case-specific" context is invaluable to the analyst who must design an appropriate estimation model, track down diverse effects, choose supporting analytical techniques, establish reasonable assumptions, and develop data that will lead to reliable calculations.

Prospective and Retrospective Studies

An economic study may be retrospective, based on empirically estimated past effects, or prospective, based on projected future effects. The longer a project has been in existence and the closer it is to market, the more feasible is an empirically based analysis. Often, by necessity, economic case studies will combine elements of existing data with forecasts in order to take the analysis into the outcomes/impact stage of a project.

Because economic case study requires impacts to be expressed in monetary units, its use is more feasible in evaluating applied research and technology development programs than basic science programs where the ultimate outcomes and impacts may be decades away and difficult or impossible to capture. However, even with applied research and technology development projects, there may be difficulties in estimation related to a project's distance from the market. Generally, the further upstream of the market a program is positioned, the more complicated becomes the task of apportioning costs and disentangling the contributions of various contributors to the development of the eventual technology, and of estimating downstream economic benefits.

Discounting Benefits and Costs

Economic case studies generally employ the techniques of benefit-cost analysis, including adjusting benefit and cost estimates for differences in their timing.

[40]The economic case study method discussed here uses microeconomic estimation techniques. Case studies may also use macroeconomic analysis models to estimate national economic effects. These techniques are discussed under the category econometric/statistical methods in Chapter 7.

Benefits and costs occurring over time are adjusted both for the real opportunity cost of capital and changes in purchasing power due to inflation or deflation in order to be compared on a consistent basis. One approach is to first eliminate the effects of inflation or deflation from the estimated cash amounts so they are expressed in constant dollars, and then apply a “real discount rate” to adjust for opportunity costs. An alternative approach is to express cash amounts in current dollars, and use a “nominal discount rate” to adjust for the combination of opportunity costs and inflation/deflation. Because an interest rate called a “discount rate” is applied to adjust the cash flows, the procedure is called “discounting cash flows.”

Discounting adjusts all dollar amounts to a common time so that they can be combined and compared with other discounted dollars. Amounts can be expressed either as a present value, a lump sum as of the present; an annual value, a series of annual amounts spread evenly over the study period; or a future value, a lump sum as of a designated future date. Often in public-sector evaluations all amounts are adjusted to present values. Discounting benefits and costs reduces the value of amounts occurring farther in the future relative to amounts occurring closer to the present time.[41]

Comparing Benefits and Costs

An evaluator must also decide how to express the measure of project performance that compares benefits against costs. Often used benefit-cost measures of project performance are briefly discussed below. All except the rate-of-return measures directly use discounting to adjust dollar amounts prior to computing the performance measure. The rate-of-return measures are computed by using

[41]The basic equation for adjusting benefits occurring in a future year t to an equivalent amount occurring at the present is Bt/(1 + d) t, where Bt = benefits in future year t, and d = a discount rate. Thus, receiving benefits valued at $1 million in five years is equivalent to receiving $712,990 today if the discount rate is 7%. And paying any more than $712,990 today for a return of $1 million in five years would be a losing proposition if a 7% annual rate of return could otherwise be obtained. An expanded set of discounting formulas, as well as multiplicative discount factors based on applying the formulas for $1.00 of value, are readily available in most benefit-cost, engineering economics, or finance textbooks to cover the various discounting operations.

the appropriate discounting formula to solve for the discount rate that equates benefits and costs.[42]

The **net benefits measure** is computed by subtracting time-adjusted costs from time-adjusted benefits. If the net benefit measure is greater than zero, then the project is considered desirable, since the minimum required rate of return is already accounted for through discounting.

When a project results primarily in cost reduction, the performance measure may be given in terms of **life-cycle costs** by combining all relevant costs and comparing them with the life-cycle costs of the best alternative to the project. A comparison of time-adjusted total costs among alternatives indicates which is lowest. If the levels of performance are comparable, the least-cost alternative is considered the cost-effective choice.

Project performance may also be expressed as a **benefit-to-cost ratio**, a variation of which is a savings-to-investment ratio. The ratio is computed by dividing benefits (or savings) by costs.[43] The ratio indicates how many dollars of benefit per dollar of cost are realized. The ratio must be greater than one to indicate a minimally worthwhile project. Again, the minimal acceptable rate of return is already built into the analysis through discounting, and a ratio greater than one means that the return is greater than the minimal acceptable rate.[44]

[42]Ruegg and Marshall provide a detailed treatment of the strengths and weaknesses of methods beginning with net benefits and ending with payback period, and demonstrate how they are calculated and used; see R. Ruegg and H. Marshall, *Building Economics: Theory and Practice* (New York: Van Nostrand Reinhold, 1990), pp. 16–104.

[43]In formulating the ratio, there are issues about which values go in the numerator and which go in the denominator. Guidance on the ratio formulation is provided by Ruegg and Marshall, Ibid., pp. 48–54.

[44]The ratio method can be used for evaluating project or program performance, but tends to be less used than net benefits for this purpose. A reason is that it does not show the dollar magnitude of net benefits, a figure normally of keen interest in evaluation. Since a ratio computed on total benefits and costs begins to fall—as a project is expanded—before the optimal size is reached, it is important to compute ratios on marginal changes when using the method to size or scope a project or to allocate a budget among competing projects.

Project performance may also be expressed as a **rate of return**. The traditional rate of return measure is an internal rate of return (IRR). This measure solves for the interest rate that will equate the stream of benefits and costs. For example, if we spent $712,990 today to receive $1 million in five years, the investment would yield a 7% internal rate of return. After the solution value of the interest rate is computed, it is compared against a specified minimum acceptable rate of return to determine the desirability of the investment or performance of a project. If we required a return of, say, 10% instead of 7%, the above investment would not be attractive.

The economics and financial communities have increasingly come to use an adjusted version of the rate of return measure that makes explicit the reinvestment rate has increased because it avoids some problems associated with use of the IRR measure, such as the possibility of no unique solution value and the assumption inherent in the technique that the rate of return on the initial investment will also be obtained on reinvested proceeds over the study period.[45]

Yet another related measure of performance is discounted **payback period**, that is, the length of time until the accumulation of time-adjusted benefits is sufficient to pay back the cost. A shortcoming of this measure is that it focuses on a breakeven point rather than on net benefits, and, hence, is not recommended as a stand-alone measure of economic performance. It may be useful, however, as a supplementary measure.

Lead with Net Benefits; Supplement with Other Measures

A frequently used strategy in benefit-cost analysis is to lead with a net benefit calculation and supplement it with one or more of the other measures to help reach audiences familiar with the different measures. Those with primarily private sector experience will usually be most familiar with IRR and business cash flows, and not the broader perspective of public sector analysis.

[45] Note that the IRR, like the benefit-to-cost ratio measure, must be applied incrementally when used for sizing projects or budget allocations.

Challenges

Challenges in expanding a case study to include benefit-cost measures are identifying the various pathways through which project effects occur, identifying the populations affected, and estimating difficult-to-quantify benefits and costs. Seldom is this information readily available. Furthermore, sufficient time may not have elapsed to allow a project to yield positive outcomes. The project may still be in the stage of net negative returns even though the potential for large positive returns in the long run may be strong. An additional challenge is attributing benefits and costs among joint investors.

Treating Uncertainty and Risk

Given the uncertainties of the technical and economic outcomes associated with research and development programs, evaluations that seek to estimate costs and benefits or other economic impacts must deal with the presence of uncertainty and risk. Quantitative studies that express results deterministically, ignoring uncertainty and risk, tend to be misleading in their implied level of precision. If probabilities can be attached to different values, risk assessment can be added to the economic analysis and the extent to which the actual outcome will likely differ from the "best guess" can be estimated.[46] If probabilities are not available, then a technique for treating uncertainty can be used. Sensitivity analysis, for instance, tests how outcomes change as the values of uncertain input data are changed, shows the estimated outcome of a project for alternative data estimates and assumptions, and allows us to express the results in terms of a range of possible values. Most importantly, it reminds the audience that there is uncertainty, and indicates how the outcome might vary. Scenario analysis allows the analyst to show results based on different scenarios of interest to decision makers.

[46]Risk assessment techniques include expected value analysis, mean-variance criterion and coefficient of variation, risk-adjusted discount rate technique, certainty equivalent technique, simulation analysis, and decision analysis. See R. Ruegg, "Risk Assessment," in R. Dorf, ed., *Engineering Economics and Management* (Boca Raton, FL: CRC Press and IEEE Press, 1996), pp. 1953–1961.

Advantages and Disadvantages of Economic Case Study

An economic case study is widely considered one of the more highly developed methods of evaluation because of its focus on ultimate outcomes and impacts rather than on outputs. Its advantages include the fact that its scope extends from project start to finish, and it provides quantitative estimates of results that are often considered more convincing evidence of value than qualitative measures. Another advantage is that its measures are stated in the language of finance, which facilitates comparisons. Combined with a descriptive treatment of a project, a well-done economic analysis can shed light on the overall performance of a project and provide valuable insight to program administrators and policy makers.

The method also has disadvantages. For instance, it may be impossible to estimate the value of important benefits in monetary terms.[47] A further problem can arise if there is not a clear understanding of the essential differences between analyses performed for public versus private projects. For instance, spillover effects in case studies of publicly funded projects designed to deliver social benefits may be overlooked in the face of easier-to-capture private returns. A related disadvantage is that stakeholders may expect positive net benefits and large IRR in the short-run when, in fact, a public R&D program often takes substantial time for impacts to be realized, particularly spillover impacts resulting from knowledge dissemination. Another disadvantage may be the risk of raising expectations based on a single project that all or most projects will be like that one, or the risk that policy makers will draw conclusions from an idiosyncratic experience. Some of these potential disadvantages, however, may be avoided through skillful execution, presentation, and interpretation of the studies.

See suggested references on economic case-study method at the end of this chapter, and examples from ATP in Chapter 6.

[47]Economists have pushed the quantification of effects far into difficult-to-measure areas, including environmental effects, aesthetics, human pain and suffering, and value of life.

Econometric/Statistical Methods

Econometrics is a branch of economics by which researchers empirically estimate economic relationships by applying mathematical models to structure the relationships, and by applying statistical methods to analyze economic data, estimate model parameters, and interpret the strength of evidence for the hypotheses examined. Thus, econometrics includes model building, estimation, hypothesis formation and testing, and extensive data analysis. The method employs many techniques from mathematics and statistics, and is used in a wide range of applications. The results are highly quantitative, with the specific units of measure dependent on the nature of the individual analysis.

Application of econometric/statistical methods requires considerable care and skill in (1) hypothesizing relationships that derive from, or correspond to, prior theoretical or programmatic concepts; (2) selecting or constructing measures for dependent and independent variables corresponding to the key concepts and relationships posited in theory; and (3) using and interpreting appropriate statistical tests.

Reflecting the complexity of the phenomena examined and the absence of perfect empirical data to use in models, Griliches and Intriligator state in their extensive reference work on econometrics,[48] the following, which captures the flavor of the method:

> There is, thus, a continuous interplay in econometrics between mathematical-theoretical modeling of economic behavior, data collection, data summarizing, model fitting, and model evaluation. Theory suggests data to be sought and examined; data availability suggests new theoretical questions and stimulates the development of new statistical methods. The examination of theories in the light of data leads to their revision. The examination of data in the light of theory leads often to new interpretations and sometimes to questions about its quality or relevance and to attempts to collect new and different data. (Vol. 1, Preface)

[48]Z. Griliches and M. Intriligator, ed., *Handbook of Econometrics,* vol. 1 (New York: Elsevier Science, 1983).

Hypotheses Testing

Hypothesis testing first makes a tentative assumption called the null hypothesis, denoted by H0. Then an alternative hypothesis, denoted Ha, is defined which states the opposite of the null hypothesis. The hypothesis-testing procedure generally uses sample data to determine whether or not H0 can be rejected. If H0 is rejected, then the statistical conclusion is that the alternative hypothesis Ha cannot be rejected. It may be postulated in Ho, for example, that the number of collaborative research ventures is unaffected by public-private partnership programs, and in Ha, the opposite.

Regression and Correlation Analysis

Another application of statistical methods in evaluation is in regression and correlation analysis to identify the relationship between a dependent variable and one or more independent variables and to measure the degree of association between variables. Regression analysis develops an estimating equation from sample data to make projections about one variable (the dependent variable, y) based on another variable (the independent variable, x). Correlation analysis measures the strength of the relationship between the variables, that is, the variability in y that is explained by x, typically measured by the correlation of determination or its square root, the coefficient of correlation.

An example of a possible relationship that might be tested is an increase in the numbers of patents by firms in a given industry and the number/amount of federal research grants received. An estimated regression equation could be used to predict the change in the number of patents given an increase in federal grants. It is important to note that neither regression nor correlation analyses alone prove cause-and-effect relationships; rather, the analyses indicate how or to what extent variables are associated with each other.

Production Function Analysis to Measure Productivity

Evaluators also use econometric analysis to estimate a production function, the mathematical expression of the technical relationship between inputs and outputs. The production function equation quantifies the output that can be obtained from combinations of inputs, assuming the most efficient available methods of production are used. The production function can be used to estimate the change in output from an additional input or the least-cost combination of productive factors that can be used to produce a given output. It can be used, for example, to examine the impact of federal funding on private-firm R&D productivity.

Macroeconomic Modeling

Macroeconomic models can help in economic forecasting and the analysis and formulation of public policy. For example, a macroeconomic model based on national input-output tables and using a set of structural equations to explain economic relationships might be used to analyze the national effects of a product innovation that decreases its supply cost. An example of a macroeconomic model is the REMI Policy Insight™ model, which is used to forecast national and regional economic effects of a wide range of policy initiatives and technological changes.[49]

Advantages and Disadvantages of Econometric/Statistical Methods

One advantage of econometric/statistical methods is that the methods significantly add to the analytical capability of evaluators. Use of these methods can contribute to an understanding of the relationships between inputs and outputs in the face of complex and imperfect data. Econometric/statistical methods can be used to produce quantitative results with detailed parameters, and, importantly, can be used to demonstrate cause-and-effect relationships.

The disadvantage to using these methods is that both the approaches and results may be difficult for the non-specialist to understand, replicate, and communicate. In addition, not all effects can be captured in these highly quantitative methods, which are imperfect and variable in how well they capture relationships between changing technical knowledge and economic and social phenomena.

See suggested references on econometric/statistical methods at the end of this chapter, and examples from ATP in Chapter 7.

---

[49]See "REMI Policy Insight, User Guide" for background on the economic theory underlying the REMI model and the REMI website (http://www.remi.com) for demonstration software.

Sociometric/Social Network Analysis

According to sociologists, the fact that economic behavior is embedded in networks of social ties has a profound impact on economic outcomes. There is an emerging awareness of the significance of social networks and their dynamics on the economic impacts of research and technology development among economists who are engaged in program evaluation.

There is growing interest in how social networks emerge, how social networks evolve, and how social networks affect economic behavior. Additionally, there is growing interest in applying methods of sociometrics and social network analysis to learn more about the spheres of influence of scientists, technologists, and innovators and the importance of their work, to identify evolving pathways of knowledge spillover, to improve the success of collaborative relationships, and to map the development and diffusion of human capital from projects.

Identify Networks of Information Sharing

Aside from tracking citations of patents and publications, how can evaluators define social networks of information sharing? One approach is to ask project participants to list several others outside their organization with whom they often share information, and also to list several others whose work they consider most important in the field of inquiry. The people they list are queried, and so forth. The multi-level communications network defined from the data can include affiliations and disciplines, can reveal paths of knowledge spillover, can show areas of influence, and can suggest the importance of the work of different people and the influence of one field on another.[50]

Another approach, called co-nomination analysis, asks researchers in a given field to nominate others whose work is similar to or most relevant to their own. Evaluators assume links exist between those co-nominated.

[50]The discussion of social networks draws on the treatment in K. Branch, M. Peffers, R. Ruegg, and R. Vallario, *The Science Manager's Resource Guide to Case Studies* (Washington, DC: U.S. Department of Energy, Office of Science, 2001).

Yet another approach to modeling scientific collaboration networks analyzes data from existing databases on co-authorship. According to Newman, scientists are connected if they have authored a paper together. This approach allows analysis of a large network without collecting primary data from the network participants. Network characteristics such as number of collaborators, degrees of separation between scientists, and clustering of the network and of disciplines are described and compared.[51]

Advantages and Disadvantages of Sociometric/Social Network Analysis

The sociometric/social network analysis methods of evaluation have distinct advantages, principal among them being that these methods bring into focus a dimension of the process of innovation/economic impact that tends to be overlooked in traditional economic analysis. If, as there is growing reason to expect, social networks are the most important element for understanding spillover effects, it behooves a public program to understand better how to model, assess, identify, and encourage formation of these networks. The methods also offer a research advantage in that they tend to require relatively modest data that can be obtained through survey, interview, or existing databases. The methods provide insight offered by an alternative perspective focusing specifically on the human and institutional dimensions of analysis.

A possible disadvantage of sociometric/social network analysis is that it remains largely unfamiliar to most economists, agency administrators, and program stakeholders. Furthermore, the resulting qualitative measures may be considered, in and of themselves, not very informative of a program's performance, particularly if the emphasis is on economic measures of impact. This problem should be lessened as economists and sociologists work together to fuse social network analysis with economic models.

See suggested references on sociometric/social network analysis methods at the end of this chapter, and examples of network analysis based on patent citation analysis in Chapter 8.

---

[51]M. E. J. Newman, "The Structure of Scientific Collaboration Networks." *Proceedings of the National Academy of Sciences,* 98(2): 404–409, 2001.

Bibliometrics: Counting, Citing, and Analyzing Content of Documents

Publications and patents constitute major outputs of research programs, and the large databases created to capture these outputs support the bibliometrics method of evaluation. As the term is used here, bibliometrics encompasses: tracking the quantity of publications and patents, analyzing citations of publications and patents, and extracting content information from documents.[52] Bibliometrics is used to assess the quantity, quality, significance, dissemination, and intellectual linkages of research, as well as to measure the progress, dynamics, and evolution of scientific disciplines.

Counting Publications and Patents

An easy output measure to track is the quantity of an organization's or project's publications and patents. The count may be normalized by research costs or some other measure of input to create an indicator of research productivity. Aggregated across a program, numbers of publications and patents per research dollar may serve as an indicator of program progress, and trends in outputs may be tracked over time. Adjustment can be made to account for quality differences in publication journals. Care should be exercised in making comparisons among organizations on the basis of their counts of publications and patents. Rates of patenting and publishing may vary for reasons other than productivity, and quality differences may not be adequately taken into account.

Citation Analysis

Tracking citations of publications and patents is useful for identifying pathways of knowledge spillovers. Citations may include publications citing other publications, patents citing publications, and patents citing other patents.

The frequency with which publications and patents are cited is also used as an indicator of quality and significance. The more other scientists cite a research

[52]Some treatments present these as three distinct methods: bibliometrics, citation analysis, and content analysis.

paper or patent, the greater its assumed relevance, impact, quality, and dissemination, other things being equal. Normalization approaches can be used to help control for quality differences in the citing journals. An example of a simple normalization approach is to hold the journal constant and compare the number of citations a given paper, or group of papers, receives against the average citation rate of all papers in the journal. A value greater than one indicates the paper, or set of papers, is more heavily cited than the average.

Who is citing publications or patents may also be of interest. Examining who is citing what can reveal where a field of research or a technology is moving, and show knowledge linkages among subject areas. For example, a public program may wish to know whether U.S.-owned or foreign-owned firms take up a technology it funded. It may wish to know if its research is supporting other fields of knowledge. Citations of research papers in patents may be of special interest to a research organization because the citations show how the program's research findings are being converted into technology and yielding economic benefits.

Citation analysis is also a useful adjunct to other evaluation methods. For example, it can facilitate historical tracing studies. In addition, citation analysis can be used to support social network analysis by investigating paper-to-paper, patent-to-patent, and patent-to-paper citations to identify potential intellectual linkages and clusters of relationships among researchers and organizations.

Content Analysis

Extracting content information is another way to use documents in evaluation. Content analysis can help evaluate the historical evolution of research funded or conducted by a particular organization, or trace the emergence of a field of knowledge from multiple sources. One approach to content analysis is co-word analysis, which uses key words to search text. The frequency of co-occurrence of the key words for a selected database of published articles depicts the evolution of ideas and concepts. A newer approach is database tomography, which avoids the need to pre-specify key words. The texts to be searched are entered into a computer database, and a computer-based algorithm extracts words and phrases that are repeated throughout the database, using the proximity of words and their frequency of co-occurrence to estimate the strength of their relationship. A more recent approach, textual data mining, goes beyond statistical methods and uses such techniques as artificial neural networks and fuzzy logic to extract content information from "data warehouses."

Visualization Tools

With both citation analysis and content analysis, the visual display of results aids comprehension of both the analyst and the audience. Special hardware and software programs, including SPIRE™ and Starlight, effectively illustrate the results of content analysis.[53]

Advantages and Disadvantages of Bibliometrics

A major advantage of bibliometric methods is that they are widely applicable to evaluation of programs with an emphasis on publishing or patenting. The methods can be used to address a variety of evaluation topics, including productivity trends, collaborative relationships, program innovation, and patterns and intensity of knowledge dissemination. Existing databases support the methods, and the methods scale easily, making it feasible and economical to apply them to large numbers of documents. The approach is relatively straightforward, and diverse audiences can understand the results. Another important advantage is that the methods do not burden those who are the subject of evaluation because data are obtained from existing databases. Some of the bibliometric methods can be applied to a program with a relatively short time lag. Finally, the objectivity associated with the methods lends them a high degree of credibility.

A disadvantage of bibliometric evaluation is that it treats only publications and patents as program outputs and ignores other outputs and long-term outcomes. Another disadvantage is that time must pass before extensive patent citations can be observed. Potential problems abound in the application of the methods. For example, counts indicate quantity of output, not quality; all publications are not of equal importance; and adjustment approaches may not adequately account for

[53]See a description of these visualization tools at the SPIRE website: http://www.pnl.gov/infoviz/spire/spire.html.

differences in quality and importance. The propensities to publish and to patent differ among organizations, technical fields, and disciplines for a variety of reasons, not just productivity differences. For example, mature technology areas can be expected to exhibit more citations than emerging technology areas. Works of poor quality may be heavily cited. Self-citations and friend-citations may artificially inflate citation rates as may patent citations provided by the patent examiner. Citing organizations may not have significant intellectual linkage. Though databases exist, the databases may be difficult to work with due to inconsistent and incomplete citations.

See suggested references on bibliometrics at the end of this chapter, and examples from ATP in Chapter 8.

Historical Tracing

The historical tracing method, or historiographic method, resembles the descriptive case study method in terms of providing an in-depth investigation in a storytelling mode. What sets it apart is its emphasis on tracing chronologically a series of interrelated developments leading from research to ultimate outcomes or from outcomes back to the factors that spawn them.

Forward Tracing

When the objective is to evaluate a given project, forward tracing, where the analyst starts with the research of interest and traces the evolution of related events from that point forward, is generally more manageable and cost-effective than backward tracing,[54] and produces a relatively complete portrayal of a project's impacts. Forward tracing enables the investigation of all known pathways leading forward from the project and contributes to a better understanding of the evolutionary processes of science and technology.

[54] If pathways from a project are unpromising, the investigation of them can be truncated with no further effort (i.e., the analysis stops).

Backward Tracing

In contrast, backward tracing, in which the analyst starts with an outcome of interest and traces backward to identify the critical developments that appear instrumental to the outcome, may or may not lead back to the project of interest. And if it does, the study may have a narrow focus that misses other effects associated with the project. For these reasons, the backward tracing approach seems more appropriate: (1) when the outcome is the central focus, (2) when a particular outcome is of known significance and the programmatic linkage is also known to exist, or (3) when the purpose is to show in a general way how significant outcomes are rooted in a certain type of programmatic funding or in work funded or conducted by certain organizations. An appeal of the approach is that the significance of the outcome is already established rather than evolving.

Following the Evolutionary Trail

The historical tracing method usually uses an interview/investigative approach to follow the evolutionary trail from one organization or researcher or development to the next. It identifies the relationships and linkages among key events, people, documents, organizations, and scientific knowledge. To identify linkages among people or organizations, it may use tools of social network analysis and citation analysis. To identify linkages among documents, it may also use the tools of citation analysis. In fact, historical tracing is often used in combination with other methods.

Advantages and Disadvantages of Historical Tracing

The historical tracing method tends to produce interesting and credible studies documenting a chain of interrelated developments, and by providing linkage all the way from inputs to outputs it may shed light on process dynamics.

A disadvantage of the approach is that evolutionary chains of events tend to be highly complex with many organizations and researchers involved, making it sometimes difficult to know the significance of apparent linkages. Evolving pathways and dead ends can stymie the forward tracing approach, while disconnects can frustrate the backward tracing approach.

See suggested references on historical tracing at the end of this chapter.

Expert Judgment

Experts are often called on to give their opinions about the quality and effectiveness of a research program. The experts generally render their verdict after reviewing written or orally presented evidence or making direct observations of activities and results.

Requirements, Logistics, and Mechanics

To provide a quality evaluation the reviewers must be highly knowledgeable about the subject and able to clearly articulate their opinions. They must be free of conflict of interest, and subject to clear, timely, and consistent process and evaluation criteria. To carry out their assessments the experts may be assembled in conferring panels or they may perform their reviews independently. They may be supported by staff to assist in data collection and report writing. They may express their opinions in terms of descriptive narratives, quality ratings (such as excellent/good/fair, high/ medium/low, or satisfactory/unsatisfactory), or as numerical scores (e.g., a number on a scale of 0–5).

Types of Expert Methods

Most federal government agencies typically use several types of expert review methods, including[55] (1) peer review, which is commonly used to make judgments about the careers of individual staff members, the value of publications, the standing of institutions, and the allocation of funds to individuals, organizations, and fields of inquiry; (2) relevance review, which is used to judge whether an agency's programs are relevant to its mission; and (3) benchmarking, which is used to evaluate the standing of an organization, program, or facility relative to another.

[55]National Academy of Sciences, Committee on Science, Engineering, and Public Policy, *Evaluating Federal Research Programs: Research and the Government Performance and Results Act,* 1999.

Advantages and Disadvantages of Expert Judgment

A principal advantage of the method lies in its practicality; that is, it provides a relatively quick, straightforward, feasible, and widely accepted approach to assessment. Another advantage is that it offers the chance for an interchange of ideas, which can lead to new perspectives.

At the same time, not much is known about the quality or accuracy of expert judgment as applied to R&D program impact assessment. It seems advisable to back up expert judgment with results from other evaluation methods and other supporting studies when attempting to assess complex phenomena.

Challenges to successful use of the method are to identify qualified reviewers, to keep reviewers free of bias and conflict of interest, and to calibrate reviewer ratings so as to render consistent judgments according to desired criteria.

See suggested references on expert judgment at the end of this chapter, and illustrations from ATP in Chapter 8.

Suggested Readings on Evaluation Methods

The following references, by no means comprehensive, are provided from the literature for those who wish to read further about the evaluation methods discussed above. Broad coverage of evaluation methods is also provided by a recent European counterpart report sponsored by the European Commission.[56]

Survey

A recommended general source of information on the survey method is the American Statistical Association, Washington, D.C., which has a section on survey research methods. *The Journal of the American Statistical Association*

[56]Gustavo Fahrenkrog, Wolfgang Polt, Jamime Rojo, Alexander Tubke, and Klaus Zinöcker, eds., *RTD Evaluation Toolbox—Assessing the Socio-Economic Impact of RTD Policies,* Strata Project HPV1CT 1999–00005, IPTS Technical Report Series (Seville, Spain: Joint Research Centre, European Commission, 2002). Additional information about the document can be found at http://www.jrc.es by searching on publications.

includes many articles on the survey method. See its website at http://www.amstat.org/sections/SRMS.

Survey analysis software is reviewed on-line at http://www.fas.harvard.edu/~stats/survey-soft/.

Commercial survey research companies are listed on-line by the Council of American Survey Research Organizations at http://www.casro.org.

A. Fink and J. Kosecoff, *How to Conduct Surveys: A Step-by-Step Guide,* 2nd ed. (Thousand Oaks, CA: Sage, 1998).

F. J. Fowler, Jr., *Survey Research Methods,* 2nd ed. (Newbury Park, CA: Sage, 1993).

D. Dillman, *Mail and Internet Surveys: The Tailored Design Method,* 2nd ed. (New York: John Wiley, 2000).

P. Krishnaiah and C. Rao, *Handbook of Statistics 6: Sampling* (New York: Elsevier, 1988). Abstract and ordering information for series available online at http://www.elsevier.nl/inca/publications/store/5/2/4/3/8/9/.

S. L. Lohr, *Sampling: Design and Analysis* (Pacific Grove, CA: Duxbury Press, 1999).

S. Sudman and N. M. Bradburn, *Asking Questions: A Practical Guide to Questionnaire Design* (San Francisco, CA: Jossey-Bass, 1982).

Case Study: Descriptive

K. M. Eisenhardt, “Building Theories from Case Study Research,” *Academy of Management Review* 14(4):532–550, 1989.

D. Roessner, “Quantitative and Qualitative Methods and Measures in the Evaluation of Research,” *Research Evaluation* 8(2):125–132, 2000.

R. K. Yin, *Case Study Research: Design and Methods,* 2nd ed., *Applied Social Research Methods Series,* vol. 5 (Thousand Oaks, CA: Sage, 1994).

K. Branch, M. Peffers, R. Ruegg, and R.Vallario, *The Science Manager's Resource Guide to Case Study* (Washington, DC: U.S. Department of Energy, Office of Science, 2001).

Case Study: Economic Estimation

The following journals regularly publish papers that help convey the breadth of program impacts that are treated by economic case study as well as other evaluation methods:

The American Journal of Evaluation, Elsevier Science Press.

Evaluation and Program Planning, Elsevier Science Press.

Journal of Technology Transfer, Technology Transfer Information Center of the National Agricultural Library. Available on-line at http://www.nal.usda.gov/ttic/JTechTransTOC.htm.

Also see the following:

A. Boardman, ed., *Cost Benefit Analysis: Concepts and Practice,* 1st ed. (Englewood Cliffs, NJ: Prentice-Hall, 1996).

P. A. David, D. Mowery, and W. E. Steinmueller, "Analyzing the Economic Payoffs of Basic Research," *Economics of Innovation and New Technology,* 2(1):73–90, 1992.

E. A. Gramlich, *Guide to Benefit-Cost Analysis,* 2nd ed. (Prospect Heights, IL: Waveland Press, 1997).

E. Mansfield, "Social Returns from R&D Findings, Methods, and Limitations," *Research Technology Management,* November/December: 24–27,1991.

R. Ruegg, "Economic Methods." In F. Keith and R. West, eds., *CRC Handbook of Energy Efficiency* (Boca Raton, FL: CRC Press, 1997).

Econometric/Statistical Methods

A five-volume reference work on econometrics provides extensive coverage of the field as it has evolved from the time the first volume was published in 1983 to the time of the latest volume: Z. Griliches and M. Intriligator, eds., vols. 1–3; R. Engle and D. McFadden, eds., vol. 4; J. Heckman and E. Learner, eds., vol. 5, *Handbook of Econometrics* (New York: Elsevier).

A four-volume reference work on mathematical economics provides extensive coverage on what may be considered background requirements for the practice of econometrics: K. Arrow and M. Intriligator, eds., vols. 1–3; W. Hildenbrand and H. Sonnenschein, eds., vol. 4, *Handbook of Mathematical Economics* (New York: Elsevier).

[The contents of the above-listed handbooks and ordering information can be found online at http://www.elsevier.nl/locate/hes.]

E. Mansfield, "Academic Research and Industrial Innovation," *Research Policy* 20(1):1–12, 1991.

Regional Economic Models, Inc., "REMI Policy Insight, User Guide." Information about the REMI Economic Forecasting and Simulation Models, including a link to consulting partners, a client list, a list of seminars and workshops, and downloadable operational demo software is available on line at http://www.remi.com.

Sociometric/Social Network Analysis Methods

International Network for Social Network Analysis. This website contains an extensive bibliography to social network analysis, research, software, courses, and other resources. Available at http://www.sfu.ca/~insna/.

Social Networks Journal. This international quarterly journal publishes papers concerned with the structure of linkages connecting "social actors." Published by Elsevier Science Publishers B. V. (North Holland) in association with the International Network for Social Network Analysis. The editorial home page for the journal is available at http://moreno.ss.uci.edu/snjhome.html.

J. Podolny and J. Baron, "Resources and Relationships: Social Networks and Mobility in the Workplace." *Working Paper Series,* Stanford University, 1999. The authors analyze the mobility among employees of a high technology firm. (View abstract and ordering information online by searching on title at http://papers.ssrn.com.)

J. Liebeskind, A. Lumerman Oliver, L. Zucker, and M. Brewer. "Social Networks, Learning, and Flexibility: Sourcing Scientific Knowledge in New Biotechnology Firms," *Working Paper Series,* Stanford University, 1995. The authors examine how new biotechnology firms use boundary-spanning social networks to increase both their learning and their flexibility. (View abstract and ordering information by searching on title or author's name at http://papers.ssrn.com.)

L. C. Freeman, "Visualizing Social Networks." Available at http://moreno.ss.uci.edu/groups.pdf.

L. Garton, C. Haythornthwaite, and B. Wellman, "Studying Online Social Networks," *Journal of Computer-Mediated Communication*, 3(1), 1997. The authors examine computer networks connecting people or organizations as a social network. Available online at http://www.ascusc.org/jcmc/vol3/issue1/garton.html.

G. Georghiou, W. Giusti, H. Cameron, and M. Gibbons, "The Use of Co-Nomination Analysis in the Evaluation of Collaborative Research." In A. F. J. Van Raan, ed., *Handbook of Quantitative Studies of Science and Technology* (New York: Elsevier, 1988).

M. E. J. Newman, "The Structure of Scientific Collaboration Networks," *Proceedings of the National Academy of Sciences,* 98(2):404–409, 2001.

Research Value Mapping Program, Institute for Policy Research and Development School of Public Policy, Georgia Institute of Technology. Barry Bozeman, RVM Program Director. A description of the research program and listing of studies is available at http://rvm.pp.gatech.edu.

J. Scott, *Social Network Analysis* (Beverly Hills, CA: Sage, 1991).

Bibliometrics: Counting, Citing, and Content Analysis

CHI Research Inc., of Haddon Heights, N.J., performs publication-to-publication citation searches, patent-to-publication searches, and patent-to-patent searches for government and other clients. Examples of bibliometric studies are posted on the company's website http://www.chiresearch.com.

The Institute for Scientific Information (ISI) in Philadelphia, PA, provides access to ISI citation databases covering thousands of international journals. ISI provides an integrated platform of networked resources for bibliometric research and offers desktop access to cited references, provides users with updates of citation information, and offers training courses in the use of its databases. See http://www.isinet.com.

U.S. Patent and Trademark Office (http://www.uspto.gov) offers online and offline search capability for patent citations.

S. Cozzens, "Literature-Based Data in Research Evaluation: A Manager's Guide to Bibliometrics." Report to the National Science Foundation, Washington, D.C., 1989.

E. Geisler, ed., *The Metrics of Science and Technology* (Westport, CT: Quorum Books, 2000). Can be ordered from the Greenwood Publishing Group at http://www.greenwood.com.

S. Popper, "Economic Approaches to Measuring the Performance and Benefits of Fundamental Science." RAND PM–409–OSTP, RAND, 1995.

C. Wagner, "Techniques and Methods for Assessing the International Standing of U.S. Science." RAND MR–706.0–OSTP, RAND, 1995.

S. Risch, D. B. Rex, S. T. Dowson, T. B. Walters, R. A. May, and B. D. Moon. "The STARLIGHT Information Visualization System." In S. Card, J. Mackinlay, and B. Shneiderman, eds., *Readings in Information Visualization* (San Francisco: Morgan Kaufmann, 1999), pp. 551–560.

SPIRE™ website: http://www.pnl.gov/infoviz/spire/spire.html.

H. Small, “A Sci-Map Case Study: Building a Map of AIDS Research,” *Scientometrics,* 30(1): 229–241, 1994.

Historical Tracing

National Science Foundation, National Cancer Institute, U.S. Department of Defense, “TRACES, Studies—Backward Tracing of Selected Major Technological Innovations to Key Events in Agency R&D Funding History (conducted in 1967, 1973, 1989, 1990).

B. Bozeman, J. Rogers, D. Roessner, H. Klein, and J. Park, “The R&D Value Mapping Project: Final Report.” Prepared for the U.S. Department of Energy, Office of Basic Energy Sciences. (Atlanta, GA: Georgia Institute of Technology, 1998). Available at http://rvm.pp.gatech.edu.

P. Kroll, G. Ault, and F. Narin, “Tracing the Influence of Basic Scientific Research on Biotechnology Patents: A Case Study,” *Patent World,* March: 38–46, 1998.

F. Narin, K. Hamilton, and D. Olivastro, “Linkage Between Agency-Supported Research and Patented Industrial Technology,” *Research Evaluation* 5:183–187, 1995.

J. Perko and F. Narin, “The Transfer of Public Science to Patented Technology: A Case Study in Agricultural Science,” *The Journal of Technology Transfer* 22(3): 65–72, 1997.

Expert Judgment

B. Bozeman, “Peer Review and Evaluation of R&D Impacts.” In B. Bozeman and J. Melkers, eds., *Evaluating R&D Impacts: Methods and Practice* (Boston: Kluwer, 1993), pp. 79–98.

Committee on Science, Engineering, and Public Policy (COSEPUP). *Evaluating Federal Research Programs: Research and the Government Performance and Results Act.* (Washington, DC: National Academy Press, 1999). Available at http://www.nap.edu/catalog/6416.html.

E. Geisler, "The Metric of Peer Review." In E. Geisler, ed., *The Metrics of Science and Technology* (Westport, CT: Quorum Books, 2000), 217–242.

R. Grol and M. Lawrence, eds., *Quality Improvement by Peer Review* (New York: Oxford University Press, 1995).

R. Kostoff, "The Principles and Practices of Peer Review," *Science and Engineering Ethics* (Special Issue), 3:19–34, 1997.

National Research Council, Committee on the Department of Energy—Office of Science and Technology's Peer Review Program, *Peer Review in Environmental Technology Development Programs*. (Washington, DC: National Academy Press, 1999). Available at http://www.nap.edu/catalog/6408.html.

CHAPTER 3

ATP's Evaluation Program

Background: Evaluation Drivers

In large measure, the development of a strong evaluation program by ATP was internally driven. ATP's standing as an experimental undertaking, established by the Omnibus Trade and Competiveness Act of 1988, together with the perspective of its first director,[57] contributed to an environment of curiosity and learning. Evaluation was seen as a tool of learning. A small amount of ATP's initial 1990 budget was set aside by program officials to fund rudimentary evaluation activities. This interest in evaluation continued to grow over time. A time line of events and developments critical to ATP and its evaluation program during its first decade is provided below:

1988 Omnibus Trade and Competitiveness Act (P.L. 100–418) authorizes ATP.

1990 ATP receives its first budget of $10 million.

1992 American Technology Preeminence Act (P.L. 102–245) amends P.L. 100–418.

1993 Government Performance and Results Act (GPRA) requires evaluation.

1993 Scale up of ATP under consideration.

1996 Progress report on ATP's impacts due to Congress.

1997 Secretary of Commerce Daley orders a 60-day review of ATP.

1998 Senate Report 105–234 requests independent assessment of ATP.

[57]ATP's first director, George Uriano, with his background combining scientific, administrative, and business expertise, played a key role in shaping ATP and its early evaluation effort. (For more information on Mr. Uriano's background, see NIST Press Release 94–36, dated September 13, 1994.)

The American Technology Preeminence Act of 1991, enacted in 1992, amended the legislation establishing ATP, and directed that:

> The Secretary [of the Department of Commerce] shall, not later than 4 years after the date of enactment of this Act, submit to each House of the Congress and the President a comprehensive report on the results of the Advanced Technology Program ... including any activities in the areas of high-resolution information systems, advanced manufacturing technology, and advanced materials.

Knowing it would have to report on results no later than 1996 provided another reason for ATP to strive to build its evaluation capabilities.

Because of these and other forces at work, ATP was able to take passage of the GPRA in stride. The GPRA's requirement for the reporting of particular types of performance metrics influenced ATP to produce those measures. Measures that ATP considered important, such as spillover measures and collaboration results, seemed to fit GPRA requirements less well, while numbers of projects funded and completed, and other accomplishments that lent themselves to counting and trend lines, seemed to fit better. Therefore, ATP established the tracking of a set of performance metrics acceptable for GPRA reporting, while continuing its in-depth, more complex economic and sociological evaluation studies that were highly informative of the workings and long-term performance of the program, though impossible to grasp in a single number or trend line.

With a scale up of ATP under consideration in 1993 and a change in the makeup of Congress in 1994, the program became the object of considerable debate and intense scrutiny. The General Accounting Office and the Office of the Inspector General carried out a number of studies of the program, and members of Congress asked the program to address many questions. In this environment, the importance of evaluation increased for ATP.

Preparation of the mandated report to Congress in 1996 also had a stimulating effect on further evaluation, providing a good opportunity to assess progress and identify shortcomings. Former Secretary of Commerce Daley's direction to the program the next year to review certain aspects of its operation gave further impetus to evaluation.

More recently, in 1998, the U.S. Senate directed ATP to arrange for a well-regarded organization with significant business and economic experience to conduct a comprehensive assessment of the program, analyzing how well it has performed against the goals established in the authorizing statute. This directive led to an extensive review of ATP from 1999 through 2001, by the National Academy of Sciences/National Research Council's Board on Science, Technology, and Economic Policy, drawing on papers presented at academy-organized workshops and roundtables and the body of work on ATP.

ATP's Evaluation Logic Model

In-depth knowledge of ATP's structure, mission, operational mechanisms, program features, and intended impacts was essential to developing its evaluation program. Fleshing out the skeletal, generic logic model of Figure 2–1 to provide a specific logic model for ATP provides a useful framework for understanding how its evaluation program was formed. Figure 3–1, supplemented by Table 3–1, depicts an evaluation logic model for ATP.

Reading Figure 3–1 from top down, ATP began with a congressional goal and approach, which, stated in broadest terms, was to increase national prosperity and quality of life by providing funding for the development of new technologies. Faced with a suite of alternative public policy strategies for supporting science and technology, Congress adopted a public-private partnership program as one element of an overall strategy to meet its goal. Congress authorized establishment of ATP, defined its mission, and provided direction to the formulation of the new program's operational mechanisms, features, and intended impacts, which were further elaborated by the U.S. Department of Commerce through the federal rule-making process. Table 3–1 lists major goals specified in ATP's mission and highlights some of the program's more important operational mechanisms and features.

ATP was directed to increase the nation's scientific and technical knowledge base, expand and accelerate development and commercialization of generic (referred to synonymously as "enabling") technologies, promote collaborative R&D, refine manufacturing processes, and increase the competitiveness of U.S. firms. It was directed to generate broad-based benefits for the nation; that is, benefits extending beyond the relatively narrow population of award-recipient

Table 3–1. ATP's Mission, Mechanisms, and Features

MISSION SPECIFICATION

- ✓ Add to the nation's scientific and technical knowledge base
- ✓ Foster expanded/accelerated technology development and commercialization by U.S. firms
- ✓ Promote collaborative R&D
- ✓ Refine manufacturing processes
- ✓ Ensure appropriate small-business participation
- ✓ Increase competitiveness of U.S. firms
- ✓ Generate broadly based benefits

OPERATIONAL MECHANISMS AND FEATURES

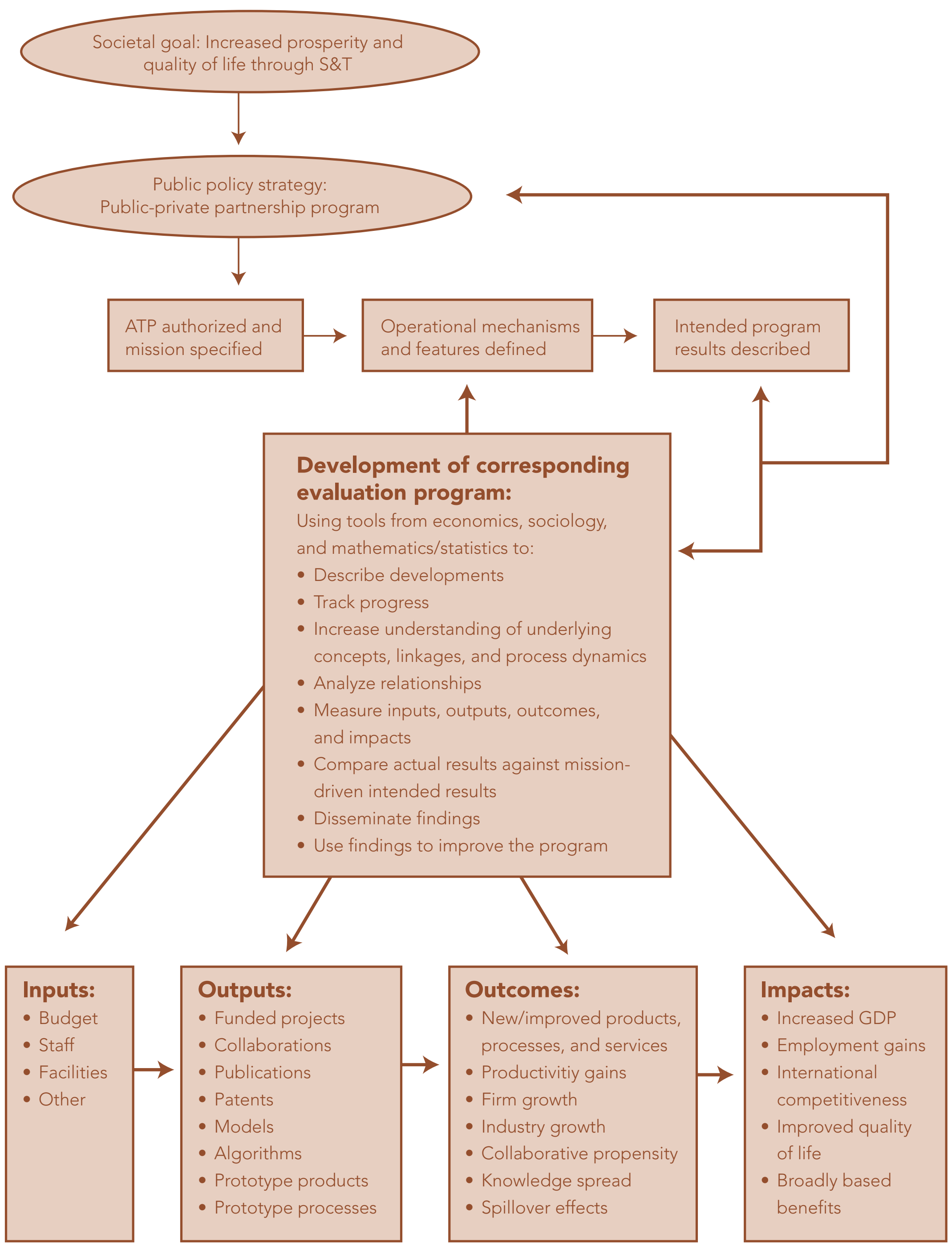
- ✓ Cooperative agreements with industry for industry-led, cost-shared research
- ✓ Focus on high-risk research to develop enabling technologies
- ✓ Competitive selection of projects using peer review and published criteria
- ✓ Sunset provisions for all funded projects
- ✓ Requirement that all project have well defined goals and identified pathways to technical and economic impacts

organizations. A constraint was included that the program should ensure the "appropriate" participation of small businesses.

Rather than using an outright, no-strings-attached grant as the award mechanism, ATP uses cooperative agreements to enter into cost-sharing arrangements with award recipients. Awarded funds can be applied only to approved costs of research. Projects are selected from proposals submitted to ATP that are peer reviewed against published selection criteria. Each award is for a specific project with both an R&D and a business/economics plan, and with well-defined goals and a limited duration.

There are two routes by which the program is designed to deliver impacts and achieve broad-based benefits: a direct route by which ATP award recipients and their collaborators accelerate development and commercialization of technologies

Figure 3–1. ATP's Evaluation Logic Model



that lead directly to private returns and market spillovers, and an indirect route by which publications, presentations, patents and other means of knowledge generation and dissemination lead to knowledge spillovers. Market and knowledge spillovers from the program are looked to as a primary means for broadening the impact of funded projects substantially beyond the direct award recipients.

The middle tier of Figure 3–1 shows the program's inputs, outputs, outcomes, and impacts. Program inputs derive from congressional appropriations that provide budgets for making awards, convening staff to administer the process, and providing for equipment, facilities, and other administrative costs. Principal outputs include the funded projects, collaborative relationships formed as a result of the program, publications, patents, models and algorithms, and prototype products and processes. Principal outcomes include sales of new and improved products, processes, and related services; productivity effects on firms; changes in firm size and industry size; a change in the propensity of firms and other organizations to collaborate; the spread of resulting knowledge through citations of publications and patents and by other means; and knowledge and market spillovers as others adopt the funded innovations. Longer-term impacts relate back to the broad societal goal that drove the program's creation, including increased GDP, employment gains, improvements in the quality of life through improvements in the nation's health, safety, and environment, and improved international competitiveness of U.S. industry. Impacts may also include an effect on the nation's capacity to innovate. Figure 3–1 also indicates "process dynamics," which refers to the transformations through which program inputs, outputs, outcomes, and impacts are linked. These transformations are complex, and there is much to learn about them.

The lower tier of Figure 3–1 ties the evaluation strategies and objectives to the program. Evaluation focuses on the inputs, outputs, outcomes, impacts, and process dynamics of a program. Evaluation objectives include tracking progress of funded projects, using, for example, indicator metrics; understanding process dynamics; estimating benefits and costs of projects and of the program overall; identifying the more difficult to measure effects; relating findings back to the program's mission; and applying tests of success (discussed in the following section). Additional objectives include disseminating evaluation results and feeding results back to program administrators to improve the program and to policy makers to inform them and to meet reporting requirements. Evaluation methods and tools used to achieve these objectives include those presented in Chapter 2.

Conceptual Tests of ATP's Success

One of ATP's central missions is to produce broad-based economic benefits. This suggests a test of success stated primarily in economic terms, and is a major factor in ATP's decision to press the use of economic methods of evaluation. However, ATP's mission is complex and multidimensional, and a single test of success is inadequate.

The following four tests can help define ATP's accomplishment of its central mission, provided they are applied after sufficient time has passed to offer a fair test:[58]

- **Test 1:** Has the portfolio of ATP-funded projects produced large net social benefits for the nation?
- **Test 2:** Has the portfolio of ATP- funded projects contributed to enhanced United States economic and technological competitiveness?
- **Test 3:** If test 1 is met, is a large share of the benefits attributable to ATP?
- **Test 4:** Regarding the distribution of net benefits, do they extend well beyond the direct ATP award recipients?

Additional criteria are needed to test for achievement of other supporting objectives while holding to program constraints. These other supporting objectives and constraints include building the scientific and technical knowledge base, fostering collaborative research, refining manufacturing, and ensuring appropriate participation of small businesses.

[58]These are expanded from suggested tests put forth by R. Ruegg, "Assessment of the ATP," in Charles W. Wessner, ed., *The Advanced Technology Program, Challenges and Opportunity* (Washington, DC: National Academy Press, 1999), p. 80.

ATP's Approach to Evaluation

Who Evaluates?

ATP's approach has been to try to capture the advantages and avoid the disadvantages of relying solely on either in-house evaluations or outside contractors. The ATP formed a core in-house group responsible for planning, guiding, and monitoring its evaluation efforts, for establishing and maintaining certain databases needed both for evaluation and program management, and for carrying out studies that could best be performed in-house.[59]

ATP has also used a variety of outside contractors to carry out evaluation studies. It formed an early association with the National Bureau of Economic Research (NBER) to ensure that it obtained the services of leading evaluators, as well as to provide another level of independent review of the studies. In addition, it formed a panel to review proposed evaluation studies. ATP periodically held a roundtable of notable evaluators who were invited in to hear and comment on presentations of studies planned and completed. ATP obtained additional reviews of evaluation studies from outside reviewers. ATP cooperated with other program assessors, including assessors from the General Accounting Office (GAO), the Office of Inspector General (OIG), and the National Academy of Sciences (NAS), to provide evaluation materials and results from surveys.

This strategy of combining in-house capability with substantial support from outside evaluators was designed to keep ATP's evaluations focused, relevant, and efficient while ensuring credibility and gaining multiple perspectives, talents, and experience.

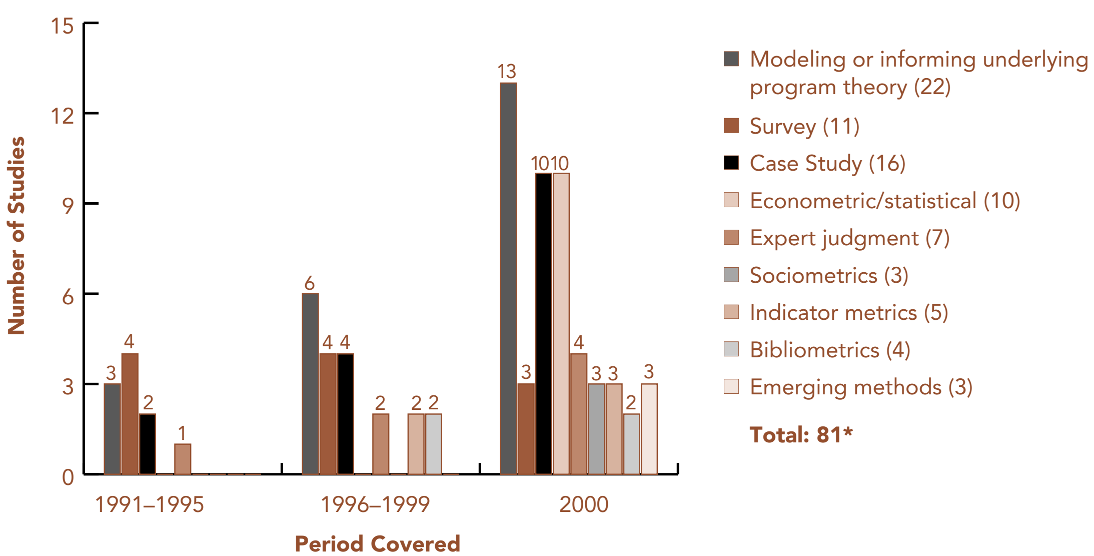
A Variety of Methods Used in a Portfolio Approach

One strength of ATP's evaluation program has been its strategy to use a variety of methods to evaluate program effects, choosing the best method for the task rather

[59]ATP's in-house evaluation staff is the Economic Assessment Office (EAO), whose primary focus is evaluation, but has other program responsibilities, such as serving on technology proposal selection boards and helping to oversee funded technology projects from an economics and business perspective.

than focusing on a single method. Taking this multi-faceted approach, ATP has given more attention to some evaluation methods than to others during its first decade. Differential emphasis on the methods resulted from programmatic considerations including mission relevancy, the time delay before certain outputs and outcomes could result, and the need to respond to specific executive or legislative requests for information. Figure 3–2 suggests the timing and relative intensity with which the various evaluation methods have been applied to ATP.

Figure 3–2. Intensity of ATP's Use of Evaluation Methods



*These 81 methods are employed in the 45 ATP studies commissioned between 1990 and 2000 that are examined in this report.

Building a Portfolio of Evaluation Studies

The sequence of evaluation studies commissioned by ATP over the past 10 years reveals how evaluation of the program evolved. Tables 3–2, 3–3, and 3–4 show the sequence of studies and indicate the principal method(s) used by each study. Two features are particularly notable: the increasing volume of studies as ATP ended its first decade and the increasing sophistication and greater focus of the studies funded. Starting with conceptual studies and surveys, ATP added case studies and econometric/statistical studies, and more recently undertook patent citation-tracing studies and network analysis.

Table 3–2. Select ATP Studies Commissioned and Completed, 1991–1995

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
1	Measuring the Economic Impact of the Advanced Technology Program: A Planning Study	1992 (Unpublished)	Albert Link, Univ. of NC-Greensboro	Program performance metrics	Informing underlying program theory
2	Advanced Technology Program: An Assessment of Short-Term Impacts—First Competition Participants	1993	Samantha Solomon, Solomon Associates	Indicators of progress toward goals	Survey
3	Estimating Social and Private Returns from Innovations Based on the Advanced Technology Program: Problems and Opportunities	1999 NIST GCR 99–780	Edwin Mansfield (deceased), Univ. of Penn.	Method of measuring private returns and market spillovers	Modeling underlying program theory
4	Economic Analysis of Research Spillovers: Implications for the Advanced Technology Program	1997 NIST GCR 97–708	Adam Jaffe, Brandeis Univ.	Market, knowledge, and network spillovers: what they are, how they arise, and how they may be deliberately pursued	Modeling underlying program theory

Table 3–2. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
5	Survey of Advanced Technology Program; 1990–1992 Awardees: Company Opinion About the ATP and Its Early Effects	1996	Bohne Silber, Silber & Associates	Indicators of progress toward goals and customer feedback	Survey
6	The ATP's Business Reporting System: A Tool for Economic Evaluation	1996	Jeanne Powell, ATP	Use of electronic survey to compile progress data from ATP participants	Survey plan
7	Advanced Technology Program Case Study: The Development of Advanced Technologies and Systems for Controlling Dimensional Variation in Automobile Body Manufacturing	1997 NIST GCR 97–709	CONSAD Research Corporation	Economic impacts of improved dimensional control in assembling vehicles resulting from a joint venture project led by the Auto Body Consortium	Economic case study; expert judgment
8	Advanced Technology Program; Early Stage Impacts of the Printed Wiring Board Research Joint Venture, Assessed at Project End	1997 NIST GCR 97–722	Albert Link, Univ. of N.C.-Greensboro	Economic impacts (mainly in terms of cost saving) of improved process technology for the Printed Wiring Board industry resulting from a joint venture project led by National Center for Manufacturing Sciences	Economic case study; survey

Table 3–3. Select ATP Studies Commissioned and Completed, 1996–1999

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
9	Acceleration of Technology Development by the Advanced Technology Program	1997 NISTIR 6047	Frances Laidlaw, ATP and G.W. Univ.	Impact of ATP on R&D cycle time	Survey
10	Development, Commercialization, and Diffusion of Enabling Technologies	1997 NISTIR 6098	Jeanne Powell, ATP	Assessment using 1995 Business Reporting System (BRS) data of progress of 480 companies and 210 projects funded 1993–1995	Survey; indicator metrics; bibliometrics
11	Small-Firm Experience in the Advanced Technology Program	1996	Jeanne Powell, ATP	Comparison of performance of small-firm awardees with all- firm awardees	Survey
12	A New Lexicon and Framework for Analyzing the Internal Structures of the U.S. Advanced Technology Program and Its Analogues Around the World	1998 *Journal of Technology Transfer* 23 (2):5–10	Connie Chang, ATP	Comparison of ATP and similar programs abroad in terms of their key features	Modeling underlying program theory
13	Advanced Technology Program's Approach to Technology Diffusion	1999 NISTIR 6385	Rosalie Ruegg, ATP	How ATP promotes early adoption/diffusion of technologies it funds by influencing project structure and firm behavior	Modeling underlying program theory

Table 3–3. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
14	Business Planning and Progress of Small Firms Engaged in Technology Development through the Advanced Technology Program	1996 NISTIR 6375	Jeanne Powell, ATP	Comparison of small-firm performance with that of medium and large firms	Survey
15	Publicly Supported Non-Defense R&D: The U.S.A.'s Advanced Technology Program	1997 *Science and Public Policy* 24(1): Feb. issue*	J-C Spender, New York Inst. of Tech.	Theoretical justification of ATP as promoting trajectories through the U.S. innovation space	Modeling underlying program theory
16	Framework for Estimating National Economic Benefits of ATP Funding of Medical Technologies	1998 NIST GCR 97–737	Sheila Martin et al., RTI	Model for estimating social, private, and public returns and application to seven tissue engineering projects	Economic case study; expert judgment
17	Papers and Proceedings of the Advanced Technology Program's International Conference on the Economic Evaluation of Technological Change:	1998 Conference date; 2001, NIST SP 952	Richard Spivack, ATP	Conference themes included public policy issues, policy goals and program design, evaluation of programs, and evaluation metrics	Modeling underlying program theory

*The paper was published independently of the ATP.

Table 3–3. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
18	Performance of Completed Projects, Status Report 1	1999 NIST SP 950–1	William Long, Business Performance Research Associates	Collection of mini-case studies of first 38 completed projects with output and outcome data compiled according to a common template	Case study; indicator data; bibliometrics; informing underlying program theory
19	Economic Impacts of Flow-Control Machining Technology: Early Applications in the Automobile Industry	1999 NISTIR 6373	Mark Ehlen, NIST	Economic impacts of adopting flow-control machining technology in vehicle production	Economic case study
20	Capital Formation and Investment in Venture Markets: Implications for the Advanced Technology Program	1999 NIST GCR 99–784	Paul Gompers and Josh Lerner, Harvard Univ.	Availability of private-sector funding for startup company R&D	Informing underlying program theory; descriptive case study
21	The Advanced Technology Program: Challenges and Opportunities	1999 NAS Press	Charles Wessner, NRC	First of 2 reports on ATP; this one summarizing deliberations of a symposium on ATP	Expert judgment informed by other methods

Table 3–4. Select ATP Studies Commissioned or Completed in 2000

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
22	Advanced Technology Program; Information Infrastructure for Healthcare Focused Program: A Brief History	2000 NISTIR 6477	Bettijoyce Lide and Richard Spivack, ATP	Genesis of ATP's Information Infrastructure for Healthcare Focused Program	Descriptive case study
23	Reinforcing Interactions between the Advanced Technology Program and State Technology Programs; vol. 1: A Guide to State Business Assistance Programs for New Technology Creation and Commercialization	2000 NIST GCR 00–788	Marsha Schachtel and Maryann Feldman, Johns Hopkins Univ.	How state programs work in combination with ATP to assist new technology creation and commercialization	Modeling underlying program theory
24	Managing Technical Risk: Understanding Private Sector Decision Making on Early Stage, Technology-Based Projects	2000 NIST GCR 00–787	Lewis Branscomb, Harvard Univ.; Kenneth Morse, MIT; Michael Roberts, Harvard Univ.	Funding gap for high-risk research	Modeling underlying program theory; expert judgment
25	Estimating Future Consumer Benefits from ATP-Funded Innovation: The Case of Digital Data Storage	2000 NIST GCR 00–790	David Austin and Molly Macauley, Resources for the Future	A quality-adjusted cost index method to estimate expected returns to investments in new technologies	Emerging method: cost-index method; econometric/statistical; economic case study

Table 3–4. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
26	Reinforcing Interactions between the Advanced Technology Program and State Technology Programs; vol. 2: Case Studies of Technology Pioneering Startup Companies and Their Use of State and Federal Programs	2000 NISTIR 6523	Maryann Feldman, Johns Hopkins Univ.; Maryellen Kelley, ATP; Joshua Schaff, New York City Democracy Network; Gabriel Farkas, Dartmouth College	Complementary use by companies of ATP, state, and other federal programs to assist them in developing technologies, and relationships among state and federal programs	Descriptive case study; informing underlying program theory
27	Advanced Technology Program's Commercialization and Business Planning Guide in the Post-Award Period	2000 NIST GCR 99–779	Jenny Servo, Dawnbreaker Press	Business planning guide to increase the likelihood of commercial success of ATP awardees in the post-award period	Modeling underlying program theory; descriptive case study
28	Development, Commercialization, and Diffusion of Enabling Technologies: Progress Report	2000 NISTIR 6491	Jeanne Powell and Karen Lellock, ATP	Assessment using 1997 BRS data of progress of 539 companies and 261 projects funded 1993–1997	Survey; indicator metrics; bibliometrics
29	Winning an Award from the Advanced Technology Program: Pursuing R&D Strategies in the Public Interest and Benefiting from a Halo Effect	2001 NISTIR 6577	Maryann Feldman, Johns Hopkins Univ.; Maryellen Kelley, ATP	Behavior of ATP award winners versus non-winners	Survey; econometrics/ statistical

Table 3–4. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
30	Performance of 50 Completed ATP Projects, Status Report 2	2001 NIST SP 950–2	ATP	Collection of 50 mini case studies, with aggregate output and outcome statistics, and project and portfolio performance scores	Case study; indicator data; bibliometrics
31	The Advanced Technology Program: Assessing Outcomes	2001 NAS Press	Charles Wessner, NRC	Report of a symposium, a collection of condensed study reports, findings, and recommendations about ATP	Expert judgment informed by other methods
32	Temporary Organizations for Collaborative R&D: Analyzing Deployment Prospects	2000 draft	Stanley Przybylinski, ERIM; Sean McAlinden, Univ. of Michigan; Dan Lura, Mich. Manufacturing Tech. Center	Estimating the propensity of a technology to diffuse	Descriptive case study; modeling underlying program theory; sociometric
33	Measuring the Impact of ATP-Funded Research Consortia on Research Productivity of Participating Firms	2002 NIST GCR 02–830	Mariko Sakakibara, UCLA; Lee Branstetter, Columbia Business School	Assessing impact of research consortia on research productivity of firms	Econometric/ statistical; informing underlying program theory
34	ATP and the U.S. Innovation System-: A Methodology for Identifying Enabling R&D Spillover Networks	2000 draft	Michael Fogarty, Case Western Univ.; Amit Sinha, Case Western Reserve Univ; Adam Jaffe, Brandeis Univ.	Identifying projects with above average spillovers	Emerging method: econometric/ social network analysis using fuzzy logic; case study

Table 3–4. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
35	The Role of Knowledge Spillovers in ATP Consortia	2000 draft	David Mowery, Univ. of California; Joanne Oxley, Univ. of Mich.; Brian Silverman, Univ. of Toronto	Internalization of knowledge spillovers among consortia members	Econometric/ statistical
36	R&D Policy in Israel: An Overview and Reassessment	2000 draft	Z. Griliches, Harvard Univ., NBER; M. Trajtenberg, Tel Aviv Univ., NBER, CIAR; H. Regev, Israel Central Bureau of Statistics	Using data from a counterpart program in Israel that had a longer operational history than ATP to demonstrate how government support of technology development fostered strong growth rates in Israel's high-tech sector	Econometric/ statistical; informing underlying program theory
37	Universities as Research Partners	2002 NIST GCR 02–829	Bronwyn Hall, Univ. of Calif.-Berkeley, NBER; Albert Link, Univ. of N.C.-Greensboro; John Scott, Dartmouth College	Contributions of universities to ATP-funded projects	Survey; econometrics/ statistical; informing underlying program theory
38	R&D Spillovers, Appropriability and R&D Intensity: A Survey-Based Approach	2000 draft	Wesley Cohen, Carnegie Mellon Univ.; John Walsh, Univ. of Illinois-Chicago	Offsetting relationship between profitability and information sharing	Econometrics/ statistical

Table 3–4. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
39	Public-Private Partnering and Innovation Performance Among U.S. Biotechnology Firms	2000 draft	Bruce Kogut, Wharton School, Univ. of Penn.; Michelle Gittelman, NYU	Increased innovation in biotech from university-firm partnerships	Econometrics/ statistical; informing underlying program theory
40	Program Design and Firm Success in the Advanced Technology Program: Project Structure and Innovation Outcomes	2002 NISTIR 6943	Michael Darby and Lynne Zucker, Univ. of Calif.-LA, NBER; Andrew Wang, ATP	How ATP promotes innovation and success of firms by encouraging collaboration and building institutional networks for cooperation	Econometrics/ statistical; social network analysis; informing underlying program theory
41	Closed Cycle Air Refrigeration Technology for Cross-Cutting Applications in Food Processing, Volatile Organic Compound Recovery and Liquefied Natural Gas Industries	2001 NIST GCR 01–819	Tom Pelsoci, Delta Research Co.	Benefit-cost analysis of a new environmentally benign industrial refrigeration for ultra-cold applications such as food processing	Economic case study
42	Study of the Management of Intellectual Property in ATP-Grantee Firms	2000 draft	Julia Liebeskind, Univ. of Southern Calif.	How conflicts over IP may inhibit success of ATP projects	Descriptive case study; informing underlying program theory
43	Determinants of Success in ATP-Sponsored R&D Joint Ventures; A Preliminary Analysis Based on 18 Automobile Manufacturing Projects	2002 NIST GCR 00–803	Jeffrey Dyer, Brigham Young Univ.; Benjamin Powell, Univ. of Penn.	Factors believed important by joint venture members to joint venture success	Descriptive case study using semi-structured interviews; informing underlying program theory

Table 3–4. (Cont'd)

STUDY COUNT	STUDY NAME	PUBLICATION YEAR AND DOCUMENT NUMBER	AUTHOR AND AFFILIATION	SUBJECT	METHOD USED (principal method listed first)
44	A Composite Performance Rating System for ATP-Funded Completed Projects	2003 NIST GCR 03–851	Rosalie Ruegg, TIA Consulting, Inc.	0–4 star rating system computed using output and outcome data from status reports to provide an overall performance measure against multiple mission goals	Emerging method: composite scoring using indicator metrics and expert judgment
45	Between Invention and Innovation: An Analysis of the Funding for Early-Stage Technology Development	2002 NIST GCR 02–841	Lewis Branscomb and Philip Auerswald, Harvard Univ.	Investigation of sources of investments into early stage technology development projects	Modeling underlying program theory; expert judgment