DOE/AL/89318--71

METHODOLOGY FOR EVALUATING ENERGY R&D

FINAL REPORT

Submitted to:

U.S. Department of Energy
Office of Science & Technology Policy (PO-81)
Project Identification Number 95-08-0013

Submitted By:

Dr. Charlie Carter
Principal Investigator
Department of Economics
Clark Atlanta University

April 23, 1997

TABLE OF CONTENTS

LIST OF TABLES

LIST OF EXHIBITS & FIGURES

ABSTRACT

Recent budgetary shortfalls and hightened concern over balancing the federal budget have placed increasing demand on federal agencies to document the cost effectiveness of the programs they manage. In fact, the 1993 Government Performance and Results Act (GPRA) requires that by 1997 each executive agency prepare a Strategic Plan that include measurable performance goals. Beginning in Fiscal Year 1999, agencies must also prepare an Annual Performance Plan that describes how actual program results will compare with performance goals. By the year 2000, the first round of Annual Reports will become due which describes actual program performance.

Despite the growing emphasis on measuring performance of government programs, the technology policy literature offers little in terms of models that program managers can implement in order to assess the cost effectiveness of the programs they manage. To date, technology evaluation literature only consist of short–term indicators of performance.

While GPRA will pose a major challenge to all federal government agencies, that challenge is particularly difficult for research–oriented agencies such as the Department of Energy. Its basic research programs provide benefits that are difficult to quantify since their values are uncertain with respect to timing, but are usually reflected in the value assigned to applied programs. The difficulty with quantifying benefits of applied programs relates to the difficulties of obtaining complete information on industries that have used DOE's supported technologies in their production processes and data on cost–savings relative to conventional technologies.

Therefore, DOE is one of several research–oriented agencies that has a special need for methods by which program offices can evaluate the broad array of applied and basic energy research programs they administer. The Office of Science and Technology Policy, which supported this project, seeks to aid DOE's program offices in their efforts to evaluate programs. More specifically, this report seeks to familiarize program offices with available methods for conducting program evaluations. To aid in that effort, this report also surveyed selected research–oriented federal agencies for possible new methods that DOE might consider for program evaluation purposes.

The general findings of the report are that few new methods are applicable for evaluation of R&D programs. It seems that peer review and bibliometrics are methods of choice for evaluating basic research programs while more quantitative approach such as ROI, cost–benefits, etc. might be followed in evaluating applied programs.

I. INTRODUCTION, BACKGROUND & PURPOSE

In the United States, concerns over the nation's economic performance has been linked to the size and composition of expenditures for research and development. The concern is not over whether the U.S. should make such investments, but how large those investments should be; how much of it should be allocated to basic research, applied research, developmental research, and other phases of knowledge production and application process; how should research resources be allocated among scientific fields, industries, and areas of national concern (defense, energy, health, the environment); and more broadly, what should be the relative roles of the public sector and the private sector in supporting our national research interests.

Decision–makers in the public and private sectors seek ways to make improvements in resource allocation decisions by improving the quantity and quality of information used to support such decisions, increasing the credibility of the rationale and justification on which they are supported, and making the decision process itself more systematic and more "rational". The greater the confidence in that process; it is reasoned, the greater the confidence that the expected outcome will be achieved. Program managers, therefore, need methods to assist in evaluating the programs before them and choose among technologies to improve confidence in the decision making process.

Altering the allocation of resources to research and development have been justified in terms of national needs (defense, regulatory support), imperfections in markets (industry tendencies to under invest in research relative to some socially optimal level), maintaining economic growth and competitiveness, improving higher education, advancing national prestige, and developing new knowledge for its own sake. Business has a narrower range of reasons for devoting resources to research, focusing on the profits expected to eventually result. Business also support research for many other reasons less directly related to profits: in maintaining state-of-the-art knowledge in certain fields, attracting and retaining scientists and engineers, ensuring an internal supply of science and engineering talent, defensive research effort, and enhancing public relations and corporate image.

Despite the disparities among reasons for supporting research, there is a common concensus that: research provides value to the supporting organization. Business and government agencies alike seek to identify the full range of values served by their research activities and, explicitly or implicitly, to use models to link research activities to the attainment of these values. It is commonly held that the more closely research efforts can be linked to measurable outcomes (that is, payoffs), the more "rational", systematic, and defensible is the resource allocation decision process. This assumption formed the basis for the Office of Technology Assessment's (OTA) comprehensive examination of the extent to which research support can be considered an investment and the returns estimated or measured (U.S. Congress, 1986).

Despite long standing interest in making R&D programs more accountable, the technology evaluation literature offers little assistance in terms of models that program managers can use to evaluate the effectiveness of R&D programs. To date, that literature has been limited to short-term measures of performance (Lambright and Rahm, 1991). Berger, et al. (1992), for instance, used the stage of commercialization of supported technologies and growth in employment of program participants as performance metrics for projects supported under the Small Business Innovation Research (SBIR) program. Other performance indicators include the amount of private or federal funds leveraged, number of firms created, licensing revenues, patents, publications and citations (Bozeman and Mellow, 1993).

Whatever measures are used to describe the range of benefits attributable to R&D programs, these measures are often compromised by absence of models that link outcome measures to supported technologies and supported technologies to technology innovation programs. In the absence of such models, it is not beyond reason to find supported technologies that could possibly have been successfully developed and applied commercially even in the absence of external support. At the extreme, it is conceivable that program participants are simply "free riders" on public R&D funds whose efforts would have succeeded even in the absence of the program's support, but were simply able to accelerate those activities or perhaps conduct the endeavor more profitably because of the program's support. In such cases the program redistributes welfare from taxpayers in general to program participants, but does not represent a net improvement in the general welfare of society.

In lieu of paucity of methods for evaluating R&D, program managers need a systematic method of evaluating technology programs they manage. Methods are also needed to assure that R&D program resources are applied in the most efficient possible way across programs so as to maximize longer–term missions and goals. The main focus of this report, therefore, is to present methodologies for evaluating R&D programs so that resources available to the U.S. Department of Energy can be used more efficiently at accomplishing its goals. Secondly, its long–term objective is to assist the agency in developing information to comply with the Government Performance and Results Act of 1993 (GPRA).

In consideration the aforementioned needs of research–oriented federal agencies, this report reviews literature and surveys R&D – oriented federal agencies for possible evaluation methods that DOE program offices could follow to conduct program evaluation. Ideally, the outcome of this effort will assist program managers in quantifying contributions energy R&D programs they manage and articulate how those technological endeavors benefit end–users– American taxpayers. The basic approach is to group energy technologies along the lines of what the science research community calls basic versus applied R&D. Methods for evaluating energy research programs are classified according to whether they are applicable to basic, applied or development research programs. Fundamental science research programs have not been subjected to as much external scrutiny as applied R&D programs due to greater difficulty in linking benefits to efforts, its smaller percentage of DOE's R&D funding and the general impression that applied technologies are appropriable. However, the need for models to support resource allocation decisions are not limited to applied and development R&D. Owing to differences in orientation of basic and applied energy R&D, energy R&D program evaluations methodologies necessarily differ.

Ideally, this report would provide DOE with a set of methods that can be implemented to evaluate the complete range of R&D programs that it administers. That objective is achieved by narrowing the unit of analysis from a larger number of technologies to more manageable set of technology innovation programs. The technology program groups must be sufficiently broad and yet be representative of the complete population of R&D technologies within the agency. More specifically, energy programs are defined according to whether the set of technologies they represent share similar objectives. The standard method of classifying R&D programs, and the approach followed in this study, is to group program units according to the phase of the effort along a continuation from exploratory–to–applied–to–developmental. Methodologies that are used to evaluate R&D programs are, therefore, grouped according to whether those methodologies are applicable to basic or applied research programs. To gain a perspective regarding the relative size of the two groups of technologies in DOE's overall R&D initiatives, cumulative spending on basic and applied programs are shown in Table 1.

The review is expected to provide a framework for evaluating basic and applied energy research and development programs. Since R&D Programs are undertaken for different purposes, methods appropriate for evaluation performance of these research efforts also differ. The report, therefore, subdivides methodologies to programs according to their appropriateness at evaluating basic energy research versus applied energy R&D Programs. Applied R&D is the largest component of energy R&D in terms of use of resources and thus constitutes the primary focus of the analysis. In terms of fundamental energy research we offer meaningful qualitative methods of evaluating performance of programs. The report also describes data limitations that hinder successful empirical implementation of the alternative approaches to program evaluation suggested in this report. Technologies have been classified according to their ultimate objectives distinguishing those with known commercial application or purpose from those that are conducted without a–priori knowledge of its ultimate end users at the time the initiative is undertaken. DOE's support of technologies of the applied or mission–oriented initiatives are called Group I technologies. Basic research which consist of research grants that enhance knowledge of the energy phenomenon without specific commercial objectives is considered Group II technologies.

Table 1

Applied and Basic Energy R&D Spending

Million of Dollars (Cumulative)

Year	Applied R&D ($)	Applied R&D (%)	Basic R&D ($)	Basic R&D (%)	Total R&D (100%)
1978	1,893.5	91.3	180.4	8.7	2,073.7
1979	4,007.6	91.2	385.1	8.8	4,392.7
1980	6,304	91.2	609.9	8.8	6,913.9
1981	8,779.5	91.3	837.4	8.7	9,616.9
1982	10,184.7	90.3	1,091.4	9.7	11,276.1
1983	11,286.1	89.1	1,386.5	10.9	12,671.6
1984	12,409.6	87.8	1,722	12.2	14,131.6
1985	13,500.7	86.4	2,132.8	13.6	15,633.5
1986	14,812.3	85.2	2,562.4	14.8	17,374.7
1987	15,976.7	83.8	3,094.4	16.2	19,071.1
1988	17,319.1	82.6	3,653.9	17.4	20,973
1989	18,707.7	81.7	4,199.8	18.3	22,907.5
1990	20,541.6	81.2	4,767.9	18.8	25,309.5
1991	22,284	80.3	5,469.7	19.7	27,753.7
1992	24,171.6	79.5	6,230.1	20.5	30,401.7

Source: U.S. Department of Energy, Energy Information Administration, Federal Energy Subsidies: Direct and Indirect Interventions In Energy Markets, Washington, DC, November, 1992.

The main body of the report is organized into Sections. A discussion of the broad R&D goals of DOE, how supported technologies relate to those goals, and the basic reasons why methodologies are needed are presented in Section I. Problems associated with R&D assessment are discussed in Section II. Program evaluations as practiced in federal agencies that also have a major research and development focus are presented in Section III. Public and private costs and benefits of energy R&D are described in Section IV. Section V describes alternative methodologies that can be used to evaluate energy R&D programs, separating those methodologies into those appropriate for basic vs. applied programs. Finally, Section V draws conclusion and recommendation regarding energy R&D program evaluation.

Energy R&D and DOE's Mission

Over the course of its 17–year history, the U.S. Department of Energy (DOE) has assumed primary responsibility for carrying out the nation energy policy. This has been accomplished through investment of public fund towards a variety of energy research and development initiatives. From fiscal year 1978 through 1993, the latest year for which actual spending data are readily available, DOE had devoted $45.5 billion to energy R&D programs aimed at developing technologies designed to achieve the broad mission included in its strategic plan. DOE was responsible for over 48 percent of domestic spending for energy research in fiscal year 1993. In fiscal year 1995, Congress appropriated $2.37 billion to DOE's fund energy R&D programs. About $1.65 billion, 9.5 percent of those appropriation went towards funding applied energy R&D programs and since the mid–1980's, about 70 percent of its total spending has gone to support applied energy R&D programs.

DOE also invests significant amounts of resources in fundamental energy research. Fundamental or basic energy research constitutes efforts directed at enhancing the nation's general understanding of energy phenomena where knowledge of specific commercial application is highly uncertain. These activities consist largely of grant–funded research projects and are usually performed by Principal Investigator at various universities. In fiscal year 1995, DOE spent $726 million on basic energy research programs in such programs as Material Sciences, Chemical Sciences, Geosciences, and Biosciences and Applied Mathematical Sciences.

Energy R&D programs contain a variety of policy initiatives designed to encourage advancement of technologies in the energy field including: financial incentives (i.e., grants and low–interest loans); regulatory interventions (including, e.g. codes and standards); demand management (e.g. through government procurement programs); and information dissemination (e.g. technology transfer networks and clearing houses of information on available technologies).

The US Department of Energy supports research aimed at discovering technologies to accomplish broad policy objectives that are of interest to the general public. Thus R&D efforts directed towards developing energy–saving technologies derives from the agency's broader goals such as energy security, environmental quality, economic strength, and scientific and technical leadership. These broader goals are discussed below and R&D activities that support to these broader goals through R&D programs. Thus technologies affect programs and it is through programs that DOE seeks to achieve its broader goals. As part of its strategic plan, DOE outlined the R&D goals for the agency. Like goals adopted by other federal R&D agencies, DOE's goals are aimed at accomplishing ends that it shares with other federal agencies. Those goals are best considered public goods and are designed to improve the general welfare of the nation since benefits accrue to general taxpayers. The characteristics of these goods are such that if left exclusively to the private sector, these needs would be inadequately addressed.

In preparation for GPRA, DOE has developed a strategic plan. In that plan it has outlined the R&D goals for the agency. Those goals are presented here to underscore the ultimate outcomes the agency needs to achieve through its R&D programs and examples of component technologies that support those goals.

U.S. Oil Vulnerability. Reduce demand-side oil vulnerabilities through research aimed technologies that offer promise at reducing vulnerability of the U.S. economy due to risks of future world oil disruptions and consequent oil price shocks.

Related R&D Activities.

- Improve the efficiency of oil use, in all sectors of the economy, with an emphasis on transportation and other oil intensive systems;
- Develop cost-effective alternatives to petroleum-derived liquid fuels, including those based on natural gas, coal, bio-fuels, and non-petroleum based electric or hybrid vehicles;
- Encourage alternative means and modes of transportation;
- Support basic research in such areas as advanced materials and underlying sciences.

World Oil Supply. Reduce supply-side oil vulnerabilities through research into technologies aimed at diversifying the world's capacity for economic production of oil, with an emphasis on domestic production, and on other production areas outside potentially unstable regions.

Related R&D Activities.

- Improve oil and gas exploration, drilling operations and reservoir characterization.
- Promote secondary and enhanced oil and gas recovery.

Energy System Resiliency. Enhance energy system resiliency through research on technologies aimed at creating a more efficient, diversified and robust (flexible, reliable) energy system, one that relies on a range of competitive and substitutable energy forms and technological choices.

Related R&D Activities.

- Improve energy efficiency in all sectors of the economy;
- Enhance diversity of energy supply, especially those of a long-term, sustainable nature, including the many forms of renewable energy, fission and fusion energy, and hydrogen;
- Improve the economic productivity of U.S. energy industries, including energy supply; energy storage; intermediate processing, transformation and refining; and distribution systems;
- Strengthen energy system reliability; reduce electric system vulnerabilities.

Air Quality. Enhance local and regional air quality through research on technologies aimed at reducing or avoiding (preventing) emissions of air pollutants from energy related sources.

Related R&D Activities

- Enhance electric power conversion efficiencies, with an emphasis on clean coal technologies, power turbine systems, and fuel cells;
- Minimize or eliminate generation of airborne waste and pollutants;

- Improve energy efficiency in all sectors of the economy with an emphasis on sources of air pollutants most adversely affecting urban and regional air quality, such as transportation and industrial technologies;

- Encourage use of non–polluting or low–polluting technologies, with emphasis on renewable energy systems; fission and fusion energy; and hydrogen;

- Improve monitoring of and quality of indoor air; and

- Enhance methods, analyses, and instruments for better understanding the longer term air quality

Greenhouse Gases. Mitigate risks associated with global climate change through research on technologies aimed at reducing, or slowing the growth in, the net global inventory of atmospheric greenhouse gases.

Related R&D Activities.

- Improve the efficiency, in all sectors of the economy, of energy–related technologies that rely on the combustion of carbon–based fossil fuels;

- Enable substitutions of lesser greenhouse gas emitting fuels and technologies for those that emit more;

- Explore energy forms that have near–zero or low net emissions of greenhouse gases, including bio–fuelsand other forms of renewable energy; fission and fusion energy; and hydrogen;

- Improve monitoring and mitigation of methane leaks and other inadvertent or controllable energy emissions of greenhouse gases;

- Enhance methods, analyses, and instruments for better understanding of global atmospherics and related environmental effects of increased greenhouse gases.

Water Quality and Land Impacts. Reduce water quality and land use impacts through research on technologies aimed at reducing the sources of impacts and improving the remedies.

Related R&D Activities.

- Reduce the contamination of surface and groundwater resources;
- Reduce, minimize or avoid the generation of waste (both toxic and non–toxic) and pollutants;
- Increase recycling, reuse, or recovery of waste products;
- Improve the recovery or detoxification of wastes;
- Mitigate natural resources conflicts and reduces energy–related land use impacts;
- Enhance methods, analyses, and instruments for better understanding the long–term environmental consequences, including those for water quality and land use, of energy production and use.

Strengthen the U.S. economy through research aimed at improving energy efficiency; creating a more productive, diversified and robust (reliable, flexible) energy supply system; promoting innovation; and encouraging competition among a range of cost–effective and substitutable energy forms and technological choices.

Related R&D Activities.

- Improve energy efficiency, in all sectors of the economy;
- Enhance the cost–effectiveness of all forms of energy supply, including renewable energy, fission and fusion energy, bio–fuels, and hydrogen;
- Improve the cost–effectiveness and productivity of energy storage; intermediate processing, transformation and refining; and distribution;
- Enhance the cost–effectiveness and environmental acceptability of energy systems;
- Reduce the economic costs of environmental compliance and improves the cost–effectiveness and management of energy–related byproducts and waste;

- Enhance methods, analyses, and instruments, for improving the reliability and comparability of data and information on energy technology availability and performance;

- Enhance international collaboration to understand better overseas requirements and gain access to markets.

Leadership in Science & Technology. Maintain and expand U.S. leadership position in science and technology by supporting a balanced and diversified research portfolio including:

- Applied research in advanced concepts and technologies across the full spectrum of energy R&D opportunities, in all sectors of the economy;

- Basic research in areas of importance to the achievement of energy–related technology objectives, including materials sciences; geosciences; energy biosciences; chemical sciences; biological and environmental sciences; super–computing and modeling; and future energy resources;

- Strategic research in multi–disciplinary fields important to the achievement of cross–cutting technological objectives, including pollution prevention; waste minimization; innovative approaches to cleaning up hazardous and toxic wastes; assessments of the impacts of global climate change; and definitions of long–term human health and environmental risks from energy production and use.

- Research investments in state–of–the–art, national, user–oriented research facilities necessary to carry out research in energy and related technical fields;

- Research investments in training and education of the next generation of scientists, engineers and technologists, who will build the base for future U.S. economic strength;

- International research collaborations, where opportunities for sharing of information and exchanging expertise within the larger global scientific community can leverage U.S. our Nation's research capabilities.

Importance of Methods to R&D Program Evaluation

Under the Government Performance and Results Act (GPRA) of 1993, the U.S. Department of Energy is required to have results–oriented management in place by the turn of the century. DOE must develop and submit strategic plans, performance plans, and report annually on how its actual performance compares with expected performance. Its purpose is "to improve the efficiency and effectiveness of energy programs by establishing a system by which the agency set goals for program performance and measure results". The Act shifts the focus by which federal agencies manage programs" from an emphasis on inputs to performances and results". GPRA list five purposes for which the act seeks to achieve, briefly stated as to:

- improve confidence among the American people in their government by holding federal agencies accountable for achieving program results.
- initiate program performance reform
- promote a new focus on results, service quality and customer satisfaction.
- help federal program managers improve service delivery
- improve congressional decision making with better information on the effectiveness of programs, and
- improve internal management of the Federal government

To achieve those purposes, GPRA calls for a consultative, interaction process of strategic planning and assessment of progress. It requires agencies to:

- develop strategic plans prior to FY 1998, consulting with congress in the process;
- prepare annual plans setting performance goals beginning with FY 99; and
- report annually to the Office of Management and Budget (OMB) on actual performance compared to goals. The first report is due in March, 2000.

The law attempts to improve program management directly through the process of producing performance goals and measures, and to improve resource allocation by taking into account performance information. It did not establish performance budgeting across the government agencies, although it did require pilot attempts in few agencies to specify the levels of results expected at different resource increments.

GPRA, requires each federal agency to submit a strategic plan to OMB by September 30, 1997. That plan is to include:

- A comprehensive mission statement
- General goals and objectives for the agency's major function;
- A description of the resources, systems, and processes that are necessary to achieves these goals;
- A description of how the general goals and objectives will be achieved; and
- A description of key external factors that could affect achievement of these general goals.

The strategic plan is also to describe how program evaluations are used in establishing goals, along with a time schedule for undertaking future evaluations. The strategic plan is to cover at least five years beyond the fiscal year in which it is submitted, and is to be updated at least every three years.

Beginning with FY 99, the Act requires federal agencies to prepare Annual Performance Plans for each program activity. A "program activity", as defined in the Act consists of "a specific activity or project" as listed in the Federal budget. The Annual Performance Plan is derived from the Strategic Plan and establishes performance goals for a fiscal year. Performance Plans for individual agencies are to be used to prepare a performance plan for the entire federal government, which is to be part of the annual budget of the United States government.

In the annual performance plan, performance goals are generally to be expressed in objective, quantifiable, and measurable units through performance indicators that measure or assess the relevant outputs, service levels, and outcomes for each program activity. The Plan must also describe the means used to verify and validate the measured values. If a performance goal cannot be expressed in an objective and quantifiable form, and alternative descriptive form may be used, but the indicators must provide a basis for comparing actual program results with the pre-established goals.

The Act established some common vocabulary for discussion of program performance. Implicitly, GPRA treats government activities and spending as inputs in a chain of activities that eventually produce benefits for the public. Government inputs are intended to produce both short-term outputs as well as longer-term outcomes.

- The act defines an **output** measure as the tabulation, calculation, or recording of activity or effort.

- An **outcome** measure, as defined in GPRA, is an assessment of the results of a program activity compared to its intended purpose.

II. STATEMENT OF PROBLEMS WITH R&D ASSESSMENT

How much of society's resources should be devoted to research? Replacing the x_i in what has become a familiar literature in welfare economics with "research" provides the theoretical answer. For a given amount of resources devoted to research, one may expect a given flow, over time, of benefits that would not have been generated had none of our resources been directed to energy research. This flow of benefits (properly discounted) may be defined as the social value of a given expenditure on research. However, if a given quantity of resources are allocated to research, this implies that these resources are not available for other activities. Therefore, by devoting resources to research, the nation is sacrificing itself of a flow of future benefits that could have been obtained had those resources been devoted to other purposes. The discounted values of the flow of benefits which we deprive ourselves of by devoting resources to research and not to other activities may be considered the social cost of a given expenditure. The difference between social value and social cost is net social value, or net social benefit. The quantity of resources that a society should allocate to basic research is that quantity which maximizes net social benefits.

Under what conditions will incentives for private profits lead to a quantity of resources allocated to research that is also socially desirable? Under what condition will it not? If all sectors of the economy are perfectly competitive, if every business firm is able to collect from society, through traditional market channels, the full value of the benefits it generates, and social costs of each business are completely reflected in the prices paid for resources used, then the invisible hand, which guides resource allocation among alternative uses generated by private profit maximizing behavior, will also direct a socially optimal allocation of resources to research.

On the other hand, when the marginal value of resources that maximizes private profits are not consistent with the social interest, the quantity will not be optimal. In such instances, the incentive for private profit maximization fail to adequately address social benefits, and, in the absence of public policy, the competitive economy will tend to devote fewer resources to research than that which would be socially desirable. Therefore, it is in the interest of society to provide some level of collective support to the production of the R&D.

Society does, in fact, collectively support a large share of the economy's overall efforts in basic research. About three-fifths of basic research effort is performed by non-profit institutions, predominantly government and university laboratories. Moreover, a large portion of basic research performed in industrial laboratories is sponsored by public funds. Much, although certainly not all, of government contribution to basic research is national defense oriented. Defense-oriented R&D aside, the American political economy recognizes basic research as an activity that creates marginal social value in excess of that which can be collected through markets. Is such treatment warranted? If so, since, in fact, society collectively sponsors much of the basic research conducted, and hence resources directed to basic research do exceed the quantity drawn by private profit opportunity, is existing social policy with respect to basic research adequate?

What are the social benefits that attend investments in basic research? It is sometimes agreed that most of the great social and political problems would simply evaporate if all citizens had a scientific understanding and, hence, that benefits derived from scientific research are only reflected in small part in the useful inventions generated by science, for science helps to make better citizens. Many scientists would argue that the very activity of sciences considered as the quest for new knowledge is itself the highest social goal and that any objective society might obtain are simply by-productsof the activity of science. Critics on both of these points are often sharp. Economists define benefits derived from the activity of sciences as the increase in value of the output flow that society ultimately receives or is made more capable of producing as a result of scientific research. In order to determine the extent to which a private firm can capture through conventional market mechanism the incremental value of output resulting from the scientific research, in particular basic research, it is necessary to examine the link between scientific research and the creation of something of economic value.

Basic scientific research may be defined as the human activity directed towards the advancement of knowledge, where knowledge consist of two types: Facts or data observed in reproducible experiments (usually, but not always, quantitative, data) and theories or relationships between facts (usually, but not always, equations). Of course, no well defined distribution can be made between basic research and other activities. Men have always experimented and observed, have always generalized and theorized. Thus, all men have been in a limited way, scientists.

Evaluating energy R&D is complicated by the wide variety of purposes behind these initiatives and by the objectives of such evaluations. For example, it may be inappropriate to apply a methodology designed to evaluate an investment by the federal government in theoretical physics as criteria to assess investments by an automobile manufacturer to increase the fuel efficiency of a new car. By the same token, methodologies used in either of the two cases would also vary but depend upon the nature and timing of the evaluation. Typically, methods and criteria used to evaluate research decisions before the investment is made (ex ante evaluation) are not exactly the same as the methods and criteria used to evaluate the performance of research investment after it has been made (ex post evaluation).

Public Good Components of Energy R&D

Approximately 90 percent of domestic energy use is derived from fossil energy. Energy generated from combustion of coal, natural gas, and coal-fired electric power plants, however, produce harmful side effects by polluting the environment. The greater the amount of energy demanded to produce goods and services, the greater the adverse effects on the environment. A major externality connected with the combustion of fossil energy is the generation of air pollution. The environment has a natural ability to assimilate a certain amount of Co_x, So_x, No_x. However, when these pollutants are generated in quantities that exceed the assimilative ability of the environment, or they are released in densely populated areas, energy use can have a detrimental effect on the quality of life. Since air quality is considered a collective good, its destruction through excessive energy use is not likely to reflect such adverse consequences or side effects. Therefore, investments into research aimed at developing technologies design to minimize unnecessary use of energy or increase reliance on energy sources that are less polluting improves the quality of the environment. The production of these technologies that makes it possible to enjoy the same quantity of material output without having to sacrifice the quality of the environment represent a public good aspect of energy R&D.

A second public aspect of energy R&D is that energy resources – coal, natural gas, and petroleum – are exhaustible. By using these natural resources excessively, future generations are deprived of their availability. The existing market forces do not adequately incorporate those future consequences in market prices of energy. Thus, unregulated use of exhaustible supplies of energy are likely to burden future generations with energy shortages. Investments in energy R&D lowers the extent to which future generations are deprived of this resource.

Energy R&D helps to protect the nation against possible adverse effects of political disturbances abroad. Approximately 40 percent of petroleum is imported from places that are capable of major political upheavals. By devoting resources to efforts aimed at reducing dependency on foreign oil, the nation's future becomes more secure. This mission is also accomplished through research whose objective is to develop alternative energy sources, enhanced discovery, extraction of domestic sources, and those that seek ways to increase use of renewable energy sources – i.e., wind–solar. Finally, by lowering the cost of energy through investments in research on technologies used in industry, the domestic industrial sector becomes more competitive relative to industries abroad. By reducing the energy intensity of domestic products, manufactured goods become more competitive in world markets and improves our balance of payments position.

In summary, there are five aspects of energy R&D that enable it to be considered a public good. First, there is a direct relationship between energy consumption and environmental pollution. By discovering technologies that conserve energy, less adverse affects on the environment occurs and the overall quality of life is, thereby, improved. Secondly, since fossil energy resources are exhaustible, excessive use by current generation burdens future generations. Since this cost of current energy use is not reflected in today's market prices, energy R&D investments that seek ways to cover this burden on future generations is a necessary public benefit that can be forthcoming only if supported by the public sector. Third, energy R&D lower the amount of energy required in production. This enhances the extent to which our industrial sector is able to compete with companies abroad. This raises exports, slows imports, and improves our net trade position with the rest of the world. Finally, a large proportion of fossil energy resources are located in politically unstable regions. Thus, the more we invest in energy R&D to develop ways to reduce that dependency, the more secure our future becomes. Again, these benefits are spread so widely that they may be considered public in nature.

Efforts by the federal government in supporting basic and applied research is so prevalent that only in recent years has that involvement come into question. Support for research in such diverse areas as energy, conservation, coal research, oil and gas exploration and extraction policy, and environmental quality antitrust, is assumed to be appropriate – if not necessary–responsibility of the public sector in general and the federal government in particular. While there always has been some disagreement over the focus of this involvement, Americans seldom argue that research in these areas ought to be conducted exclusively by the private sector.

The appropriateness of public sector involvement in R&D is fundamentally an economic issue in which research is viewed as a production process whose output is considered a public good. According to Arrow (1962), perfect competition and reliance on the private market place is not the "best" (socially optimal) way to allocate information because of three features frequently associated with the production of information: 1) indivisibilities, 2) inappropriability and 3) uncertainty. A discussion of how each of these features of the R&D process can be used to justify public sector investment in R&D follows.

The term "indivisibilities" as used in this context, applies to situations in which the scale of the activity cannot be increased or decreased in small increments. In some cases the production of information involves the use of large amounts of physical capital, for example particle accelerators and light water reactors, which, for either scientific or engineering reasons, have a single most efficient size. In these cases, it may not be scientifically prudent or technically feasible to make the equipment a different size, or the costs of production (of information) associated with other equipment sizes may be higher. In this situation the total cost of producing information increases less than proportionally with output and the long–run average cost of producing information also decreases with output.

Indivisibilities in production can give rise to situations in which competition between firms will not lead to a socially optimal allocation of resources. Indivisibilities constitute a particularly serious problem in resource allocation when decreasing average costs occurs over a wide range of output, large enough to satisfy the entire market for a particular kind of technology. In this situation, a single firm can produce the technology more cheaply than two or more firms. The type of firms that produces such output is referred to as "natural monopolies." Natural monopolies, like other forms of monopoly, are able to charge higher prices at lower levels of output than would occur under perfect competition. More importantly, the benefits that society enjoys when a natural monopoly produces the technology are not as great as the benefits that could be created when the technology is produced and distributed more widely under conditions of perfect competition. Consequently, it is argued that public sector involvement is appropriate to correct for the effects of the market distortions created by indivisibilities and, thus, to ensure that the level of production is socially optimal.

Another aspect of Energy R&D that makes it a public good is inappropriability, i.e., benefits produced from the R&D process cannot be fully appropriated by those who produce it. According to Arrow (1962) the problem lies in the fact that, while the cost of producing new information is often very expensive, the producer cannot, without special legal protection, restrict access to this information and expect to reap the full extent of the social benefits that it creates. This is because any user can reproduce the information at little or no expense and pass it along to others. Under these conditions, the optimal strategy for an agent is to become a "free rider." That is, it is to the agent's advantage to simply wait until other agents produce the information and then acquires it at a much lower cost – a situation that leads to underinvestment in new information.

Arrow argues that this problem can be addressed by establishing legal protection through the patent – copyright system and various types of licensing–royalty schemes. However, no amount of legal protection can make so intangible a product as information a thoroughly appropriable commodity. In fact, complete protection could become so costly so as to ultimately grant monopoly states to the owner of information. The owner would then be able to appropriate all of the potential benefits created by the information, with the consequence that users of this information would have to pay higher prices for less information. In short, the information would be under–utilized.

The problem of appropriating benefits from R&D applies not only to information as a commodity, but also to areas where property rights to goods and services are, for theoretical or political reasons, not well defined. The case of a public good, such as national defense, represents a broad area in which government tends to perform a leading, but not exclusive, role in R&D funding. Environmental research is an area in which ambiguously defined property rights tend to create private incentives for firms to overuse the waste assimilation capacity of the environment and to under–invest in information concerning the effects of their actions on the environment.

A final public aspect of R&D is uncertainty. The output of R&D, particularly basic research, cannot always be predicted in advance. This uncertainty can be reduced through futures markets, which reduce the risks to producers by diffusing this risk over a large number of buyers and sellers. Insurance performs a similar function. However, as Arrow (1962) points out, shifting of risks in the real world is incomplete. Under these conditions, one would expect underinvestment in risky activities and the magnitude of this underinvestment would increase with the level of risk. Since government expenditures on R&D are paid for through taxes, public sector investment in risky activities has the positive effect of diffusing risk much more widely than would be expected by private market arrangements.

In summary, then, competitive market arrangements can be expected to result in underinvestment in R&D because information is frequently subject to indivisibilities in production, because the results of R&D are difficult to appropriate, and because R&D is an inherently risky exercise. The underinvestment in R&D will tend to be greatest in basic research, where these three circumstances tend to be most accentuated. Finally, even if a firm is able to collect all of the benefits derived from an R&D investment, that information will tend to be monopolized by the firm, priced beyond the means of many prospective users, and therefore, under–utilizedby society.

Measurement

Measuring benefits and costs energy R&D is made difficult because of the indirect efforts through which basic and applied research affects the economy. Once the effects of energy R&D

on individual economic agents have been identified, the first step consist of estimating demand and supply functions for the relevant economic agents in the appropriate markets or non–market contexts. The second step consists of using these demand and supply functions to simulate the behavior of buyers and sellers, with and without energy R&D for a period of time appropriate to the specific case. Third, the results of the simulations are used to calculate the difference in the sum of producer and consumer surpluses due to energy R&D in each period. These surplus changes are then discounted in each period back to the date of origin and then summed to obtain a measure of the present value of the net benefits to society as a result of energy R&D. The present value of the energy technology's cost is calculated and subtracted from the present value of its benefits to obtain a measure of the net present value of the R&D society.

The execution of these steps is sometimes problematic due to paucity of information on measures of costs and benefits of a technology. In cases involving ex ante and, even ex post evaluations of basic research, the problems of predicting long–term consequences of energy R&D efforts and identifying markets for the information produced by that R&D make it virtually impossible to construct demand functions for that information, except perhaps for scientists and others who value this research for its own sake – and this has never been attempted. Given these measurement problems, it seems unlikely that standard economic tools offer a practical methodology for evaluating basic research projects, unless the effects of the project can be defined with enough precision to construct demand or supply curves, as required.

Construction of supply and demand curves is less difficult in cases where the information produced by energy R&D could influence, or actually has influenced, the production or consumption of market goods and services. Constructing market demand and supply curves is conceptually straightforward in cases where the major results of energy R&D have been to reduce the costs of producing an existing market good or service. In this case, the analysis of net benefits is consistent, conceptually, with the movement of supply curves. Unfortunately, constructing demand and supply curves may be limited by the proprietary nature of sales information in an industry or by prices which understate true social value of a good or service due to market distortions. The case of goods and services provided by DOE R&D contractors may represent the best example of problems caused by market distortions.

In other cases, where energy R&D has resulted in the production of a new good or the improvement of an existing good, different supply and demand curves must be constructed to reflect these changes. However, the data requirements associated with modeling the effects of quality changes on supply and demand curves are extensive and, in many cases, probably exceed the availability of information needed to conduct this type of analysis. In cases where lack of data makes it difficult to construct market supply and demand curves for the goods in question, economists may still be able to use available market data in conjunction with simplifying assumptions about the curvature of supply and demand curves to approximate changes in producer and consumer surplus attributable to the effects of R&D.

As previously mentioned, one of the major problems associated with measuring the benefits of federal R&D is that there may be no market in which to value some of the potential or actual effects of R&D. In these cases, construction of supply and demand curves for non-market goods and services has, until recently, been extremely difficult. Two traditional approaches to this problem have involved valuing these non-market effects as a residual, after the returns to all other inputs have been calculated, or else by valuing them based on the cost of inputs used to produce the effects in question. The human capital approach to valuation of non-market effects is an example of this latter approach. However, cost-based definitions of non-market values are not consistent with willingness-to-pay concepts and are particularly problematic in cases where the federal government is the only buyer in a particular market, such as in defense or space-related contracting.

More recently, two alternative approaches for measuring the benefits associated with the production and consumption of non-market goods have gained increasing acceptance. The first such method uses surveys to determine how individuals think they would behave in hypothetical situations. This approach is used to determine how much an individual would be willing to pay for another unit of a non-market good. This information is then used to construct demand curves for the non-market good. The chief strengths of this approach are that it is well-grounded in economic theory and very flexible in its application. On the other hand, the values elicited by this approach are potentially subject to a number of biases, which has made it extremely controversial.

The second of these approaches uses changes in market values - either the wage compensation of individuals or the value of property - to measure non-market effects. The major advantage of this approach is that it relies on existing market information to estimate labor supply curves or property bid and offer curves, as relevant. This approach offers an important advantage over the former method which asks people what they would spend in a hypothetical situation, but does not require them to part with their money. The main weakness of this approach is that it is less consistent with economic theory and requires fairly restrictive assumptions about the structure of relevant property markets and the relationship between property values and the non–market effects in question. In spite of their limitations, both of the approaches represent a substantial improvement over traditional approaches for measuring the benefits and costs associated with non–market effects.

A final problem area associated with the measurement of the benefits and costs of R&D involves the practice of discounting future monetary sums into present values. Two arguments are advanced to justify this practice. First, resources that are not used for immediate consumption can be employed in investment projects yielding a return in later periods. And second, society may regard consumption by future generations as somewhat more or less important than that of the present generation. The first argument generally supports the use of discount rates on federal investments which reflect rates of return on displaced resources in the private sector. The second argument is generally used to support lower discount rates to ensure that more wealth is passed along to future generations.

In short, there is no single approach to discounting, nor any single discount rate on which all economists and decision makers would agree. However, different discount rates can have a profound effect on the net present value calculated for a specific project. In general, higher discount rates make future costs and benefits worth less and tend to favor projects with immediate payoffs. As such, high discount rates would tend to hurt the relative standing of basic research investments vis–a–vis R&D investments that have near–term market applications.

III. REVIEW OF PROGRAM EVALUATION IN FEDERAL AGENCIES

U.S. research agencies generally follow one of two approaches to R&D Program evaluation. One is technical review by a panel of external experts, always including researchers and sometimes including users of research results as well. For example, since the 1950s, the National Institute of Standards and Technology (NIST) has evaluated its programs with extensive site visits by expert panels organized by a Board of Assessment, a branch of the National Research Council. In the same spirit, the Office of Energy Research of the DOE conducts a highly structured retrospective evaluation of expert assessment at the project level, with panel scoring projects based upon on pre-set criteria. The scores are aggregated at program level and reported within the agency.

A second approach to research program evaluation relies more extensively on data gathering by external contractors. Such evaluation studies, which draw more directly on the general program evaluation tradition, often use surveys or publication-based indicators, sometimes in combination with expert judgement of various sorts. An example is the National Science Foundation's survey of participants in its Research Experiences for Undergraduates Program and the National Institute of Dental Research study of restorative dental materials research. Both approaches used publication based indicators, patent indicators, surveys, and case studies.

Evaluation studies, however, are relatively rare. Those that have been done, however, are concentrated in the fundamental science agencies, NSF and NIH. Most assessments of fundamental research programs are descriptive, and far removed from the sort of quantification of performance GPRA is seeking. A large array of quantitative tools for evaluation has been described in the literature; but few of them are used in practice because they provide too narrow a view of the productivity of R&D. To respond to GPRA, research programs and agencies thus are then challenged by having to choose among a limited array of procedures they have largely avoided in the past, or developing new ones. This section reviews the experiences and direction that other federal agencies seem to be taking in R&D program evaluation. The objective is to determine if their experiences can be helpful in carrying out program evaluation efforts by DOE.

Army Research Laboratory (ARL)

The Army Research Laboratory is a relatively new organization among research–oriented federal agencies. In 1993, it was established by the consolidation of seven formerly independent Army laboratories. Its purpose is to provide the fundamental and applied research from which the future material for the nation's land warfare is expected to emerge. In its strategic plan ARL adopted the following goal: providing the Army with key technologies and analytical support necessary to assure supremacy in future warfare. The laboratory uses the peer review system to rate the six mission areas, metrics to guide the laboratory in the desired direction, and customer feedback questionnaire to judge the level of satisfaction with its deliverables.

The ARL has developed what public federal research agencies consider a rational approach to R&D evaluation . ARL first reorganized the laboratories from a focus based upon the laboratories concept to mission areas. The ARL developed six mission areas and evaluates each area in order to assess performance of the laboratory.

Mission areas are evaluated by the following methods: peer review, user surveys and metrics; and specifically addresses the issues of: productivity, relevance and quality. The mission areas for ARL are as follows:

Digitization and Communication Science: to provide the fundamental science necessary to assist developers in exploring the information technology explosion, closing the gap that exist between military and commercial information systems, and digitizations the army.

Armor/Armaments: enhancement of technologies to increase lethality and survivability of army weapons systems to maintain a qualitative edge on Future battlefields.

Soldier System: help others to assure that the soldier can operate effectively on the high-technology battlefield and survive in its lethal environment.

Air and Ground Vehicle Technology: develop the technologies needed to help others to extend the life of current combat vehicles and provide technologies for future systems.

Survivability and Lethality Analysis: provide vulnerability, lethality, and survivability assessments of fielded and developmental Army weapons systems and develop the tools necessary for efficient assessments that produces authoritative results.

Recognizing lack of a systematic method for evaluating R&D programs as the greatest problem, ARL developed a method to address the question of how well the laboratory was performing and its degree of health. The construct, as it is called, follows a semi-quantitative approach to program evaluation in that it requires the Director to take quantitative and qualitative factors into consideration in developing an assessment of the laboratory. The evaluations construct is aimed at evaluating projects according to their areas: relevance, productivity, and quality (See Table 2).

relevance: is the work being performed in response to same bonafied requirement of a customer;

productivity: on any given project, or for the laboratory as a whole, is progress being made towards some specified goal at an acceptable rate;

quality: Is the work being performed at a level that would be considered at or beyond the state–of–the–art.

The instrument is supported by three pillars, which, in various combinations, are designed to assure that the three areas of interest are addressed. There three pillars represent the primary method used by the Army Research Laboratory to evaluate its directorates. They consist of three widely used approaches to evaluation of R&D programs: Peer Review, Metrics, and Customer feedbacks.

Peer Review: The ARL entered into agreement with the National Research Council to assemble a Technical Assessment Board (TAB) to oversee its peer review process. The TAB is made up of approximately a dozen individuals with international reputations. Working under this Board are six panels with memberships ranging in size from 5 to 15 persons per panel, who also are of high status in the technical community. These panels provide TAB with a comprehensive analysis of the performance by various mission areas over a two–year period. They report findings annually to TAB which then prepares a written report that describes the performance of the Laboratory which is to be published by the NRC. The NRC has responsibility for hiring members of TAB and panels and appoints a full–time Staff Director and necessary clerical support to carry out its function.

Metrics: Although ARL is proceeding with the collection of data to measure efforts towards research, it cautions against excessive emphasis on these measures. In all, the laboratory is constructing a database that consisted of some 57 metrics, or "business drivers". Considering 57 metrics as far too many to monitor on a regular basis, the laboratory reduced the number to

26. This 26 metrics included both output and input measures. Which are used by the laboratory director to serve as "levers".

Recognizing their limitations, ARL assembled 57 metrics for consideration. Most were input metrics, but some were output measures. Metrics are classified into seven categories.

- Technical accomplishments – Count of items delivered and programs completed as well as papers, patents, citations, etc.

- technology transfer – counts of cooperative R&D agreements (CRADAS), patents licensing agreements (PLAS), visits to industry for program reviews, etc.

- esteem factors – counts of significant awards, prestigious posts held by senior researchers, invited presentations at gatherings of national societies, and the like

- fiscal performance – obligation and disbursement rates, overhead rates, in–house/ outhouse ratios

- facilities and equipment – the degree to which the lab is a state–of–the–art facility.

- personal – collection of items descriptive of this personnel structure

- greening of the workforce – a measure of the efforts made to acquaint entry– and mid–level civilian leave staff members with the "real" army

Table 2

The ARL Performance Evaluation Construct

Performance Measurement	Why ...		
How ...	Relevance	Productivity	Quality
• Peer review	○	✪	●
• Metrics	✪	✪	✪
• Customer evaluation	●	●	✪

● = Very Useful ✪ = Somewhat Useful ○ = Less Useful

Source: Dr. Edward A. Brown, "Conforming the Government R&D Function with the Requirements of the Government Performance and Results Act: Planning the Unplannable? Measuring the Unmeasurable? Measuring the Unmeasurable? (Unpublished manuscript)

Table 3

ARL Uses Peer Review to Evaluate Mission Areas

Performance Evaluation Category	Digitization and Communication Science	Armor and Armaments	Soldier Systems	Air and Ground Vehicle Technology	Survivability/ Lethality Analysis
1. Technical Program	---	---	---	---	---
2. Quality of Staff	---	---	---	---	---
3. State-of-the-Art Equipment	---	---	---	---	---
4. State-of-the-Art Facility	---	---	---	---	---
5. Relevance of Work	---	---	---	---	---
TOTAL					

Source:

Each cell is assigned a score value ranging from 1 to 5 (Poor -to -Outstanding). The rating is actually carried out by Directorate, but since Directorates are aggregated to form mission areas summary scores by directorate enable the ARL to develop summary ratings by mission areas. Then the mission summaries are aggregated to develop an assessment for the laboratory.

S

Although data is collected to develop on 57 metrics, that number was considered too many to deal with on a regular basis. The Director dealt with 26 of the 57 that were of special importance, and which serves as levers to move the organization in a desired direction. These 26 metrics used by the lab are reported in table – and classified according to whether they represent input as output proxies. These metrics are listed in Table 3.

Customer Feedback: third pillar in ARL's R&D evaluation procedure consist of a short questionnaire sent to selected groups of customers. It considers customers for the immediate short–run products of its efforts. These were defined as Research Development and Engineering Centers (RDEC) and the various system Program Managers and Program Executive Officers (PMs/PEOs). Aside from these internal stakeholders, ARL also surveys the end item users – the soldiers.

ARL's mission areas are evaluated separately and aggregated to arrive at an overall assessment for the laboratory. Categories used in determining the performance of the mission areas are: Technical Program, Quality of Staff, Quality of Equipment, Status of Arts Facility, and Relevance of Work. ARL's plans to evaluate each of its five mission areas according to the five categories using the NRC to oversee panels of external experts. Thus, the Lab method of program evaluation might be thought of as a five by six matrix whose rows represent attribute and columns missions areas. The elements of the matrix consist of scores that external review teams have assigned to five categories by mission area. Table 2 outlines the basic tenants of this measurement matrix.

Table 4

ARL LIST OF PERFORMANCE METRICS

Output Measures	Input Measures
Percentage of "Top 5" tasks completed	Glidepath Annual Manpower Target
Percentage of Focus Programs completed	Percentage of Ph.D.s among the Scientist and Engineer (S&E) portion of the workforce
Percentage of Science & Technology Objectives (STOs) completed	
Customer Satisfaction Rating (from customers whom are not charged)	Percentage of employees with 40 + hours of training per year
Customer Satisfaction Rating (from reimbursable customers)	Number of employees on long–term training
Number of invention disclosures	Funding from customer sources
Number of refereed journal articles	Obligation rate
Number of ARL technical reports	Disbursement rate
Number of Cooperative R&D Agreements (CRDAs) established during the fiscal year	Indirect overhead rate
Number of Patent Licensing Agreements (PLAs) established during the fiscal year	Percentage of employees completing the basic ARL "Greening" course
	Number of employees completing the FAST, Jr. training course
Number of post–doctoral fellows hosted at ARL	
Number of post–doctoral mentors on the ARL staff approved by the National Research Council	
Number of guest researchers at ARL	
Number staying more than three months	
Number of guest researchers from ARL	

Source: Dr. Edward A. Brown, "Conforming the Government R&D Function with the Requirements of the Government Performance and Results Act: Planning the Unplannable? Measuring the Unmeasurable? Measuring the Unmeasurable? (Unpublished manuscript)

Table 5

PERFORMANCE ASSESSMENT OF DIRECTORATE BY "TOP 5 DELIVERABLES"

Directorate	Five Most Important Tasks				
	One	Two	Three	Four	Five
Digitization & Comm. Science	Green	Green	Yellow	Yellow	Green
Armor & Armaments	Yellow	Yellow	Red	Red	Green
Soldier Systems	Green	Red	Green	Green	Red
Air & Ground – Vehicle Technology	Green	Yellow	Green	Yellow	Green
Survivability/Lethality Analysis	Yellow	Red	Red	Yellow	Red

Source: See Table 4.

To make the most efficient use of its data on metrics, ARL developed a system to quantify achievement by its directorates with respect to performance on the top 5 deliverables (Table 5). For each of the five most important tasks, each directorate was given a three color (Red, Yellow, and Green) based upon the extent to which it had achieved the specific task. Each color was assigned a number (Red = 0, Yellow = 0.5, and Green = 1). Thus, a directorate considered to be performing at the highest standard would receive a score of 1 while one performing unsatisfactory would secure a score of 0. The raw score was measured relative to the highest possible score. For instance, suppose that the Armor and Armaments directorate is judged in its top 5 task as: GGYYG = 3G + 2Y. With G = 1 and Y = .5, then the Directorate would be given a raw score of 4 = 3(1) + 2(.5). Its percentage score would, therefore, be 4/5 = 80%. Evaluating each directorate accordingly enables the laboratory to derive an overall performance assessment for the laboratory.

Information gathered through customer feedback also is used in a way that quantifies performance of the laboratory's efforts to satisfy customers. The survey is included in the appendix, but question areas and hypothetical responses are provided here to illustrate its use in performance evaluation. The responses to each issue/area are allowed to vary from poor to excellent. Once the results are collected, they can be easily described to provide a general overview of the level of satisfaction of its customers to work done by the laboratory.

Summary measures from the peer review system can be easily derived to represent the current health of the LAB. Contents of the 6x5 matrix can be crosstabulated by mission areas or by performance category to highlight the health of the lab.

Table 6

HYPOTHETICAL RESULTS OF USER EVALUATION SURVEY

Directorates	Technical Quality (x_1)	Value Received (x_2)	Utility of Product (x_3)	Working Relationship (x_4)	Overall Performance (x_5)
1	Poor	Good	Good	Poor	Good
2	Fair	Poor	Good	Poor	Poor
3	Fair	Fair	Good	Good	Fair
4	Good	Good	Poor	Fair	Good
5	Good	Excellent	Very Good	Good	Fair
6	Excellent	Fair	Excellent	Fair	Very Good

	Poor	Fair	Good	Very Good	Excellent
Technical Quality	1	2	2	0	1
Value Received	1	2	2	0	1
Utility Product	1	0	3	1	1
Working Relationship	2	2	1	0	0
Overall Performance	1	2	2	1	0

National Science Foundation

In 1995, the National Science Foundation devoted $3.3 billion to support education projects in the sciences, engineering, and social sciences. Although the Foundation manages these research projects, the actual research is performed almost exclusively at Universities, the mission that the Foundation was given in 1950, the year it was established. Along with DOE and other federal research agencies, NSF has been developing methods to evaluate its enormous basic research programs.

NSF has already developed the first document that is required under the GPRA: A Strategic Plan. The plan articulates three broad goals for the foundation and provides four core strategies for achieving those goals. NSF's three goals are: (1) to enable the U.S. to have world class science, engineering, and mathematics; (2) to place new knowledge in service to society; and (3) to achieve excellence in science education.

The foundation provides the following core strategies designed to accomplish the above three goals:

(1) Develop intellectual capital
(2) Build the physical infrastructure
(3) Integrate research and education
(4) Promote partnerships

Through pilot projects and internal concept papers, NSF has moved forward by experimenting with methods to transform its strategic goals into performance objectives and to develop indicators of the four types of NSF initiatives: research; education; facilities; and administration and management. Performance indicators are being developed in all four areas.

One drawback is that the indicators are not easily aggregated into the performance concepts that are expressed in the strategic plan.

NSF is experimenting with a portfolio approach to evaluate performance of the facilities, centers, and project grants. The Foundation supports a number of user facilities, such as telescope and accelerator, in several of its Directorate. These facilities share the Foundation's common goal as phrased in its strategic plan, "to enable the United States to uphold a position of world leadership in all fields of science." Each of the Foundation's facilities was established in response to needs in specific fields, and each began from a different technical baseline. Relying on information supplied by facility users, a dozen generic performance indicators were developed for these facilities.

The key element necessary in applying the model is to represent performance measures in terms of percentage change from a baseline. The baseline, number or index could differ from one facility to another, and even be measured in different units. To standardize, NSF developed a concept called "user units" to refer to measurements such as team time and observing hours. In the end, the percentage change from each facility could be aggregated into an overall index to characterize performance for the entire portfolio of facilities. NSF prefers the portfolio approach because it minimizes the consequences of large variation in performance of particular facilities on the overall index for the portfolio. Over time, the index should improve as old facilities are taken off line and new more efficient ones installed.

For other basic research investments, NSF established goals for the Foundation by following the descriptive format allowed under GPRA. Plans are to summon panels of external experts and use available performance indicators along with the wisdom and experience of members of these panels to judge performance of its Directorates in relation to the Foundation's goals.

NSF considers itself as a research managing agency and is developing metrics to identify efficiency measures for such innovations as electronic submission of proposals. Projects are selected under the condition that they offer the greatest promise of innovating and, thereby, changing the structure and problems of bodies of fundamental knowledge. These bodies of knowledge are used by different problem–solving organizations and their ever–changing character is important to establishing value. Likewise, the research process maintain the skill and knowledge of researchers and the students trained. The human capital of the nation is thus improved through the research process, to be available for many different uses. NSF justifies its supports of a diversified portfolio of projects because of the difficulty in predicting the field from which the next major innovation will come.

There is a broad concensus on the most common, tangible product of the research process and many research organizations rely on them in evaluating research. New ideas are of little value to society unless those ideas are communicated to larger audiences. Students are major modes by which ideas and technical competence are transported into the economy. Faculty status is largely based upon scholarly activity, and publication of research findings is viewed favorably by peer reviews in the project selection process. Other tangible outputs of the basic research process – i.e., undergraduate teaching, inventions, collaborations with researchers from problem solving organizations are also among the useful measures of basic research output.

NSF also considers support of graduate student educational training as a way of maintaining the nation's intellectual capital. Likewise, the Foundation considers publications as an indication of the extent to which findings through basic research it sponsors are shared. NSF views its role as guardian of the knowledge base and promoting the effective use of that knowledge base. NSF considers the value of fundamental research it supports to be reflected in its influence on the economy and the quality of life of American citizens. Output of basic research projects represent intermediate goods in the sense that those outputs represent work–in–progress and, therefore, must be further fabricated and combined with other outputs before they are in a usable form before their impact on the economy and quality of life of American citizens is recognized.

In recognition of the long gestation period before investment in basic research produce recognizable outcomes, it is impossible to set performance goal for such abstract measures as levels of influence on the economy or on society at large on an a–priori basis. Under the GPRA a performance goal is considered a target level of performance expressed as a tangible, measurable objective, against which actual achievement can be compared.

NSF's strategies goals are: to building technical capacity for the nation and is accomplished through form core strategies: developing intellectual capital, building the physical infrastructure, integrating research into teaching, and promoting partnerships. NSF relies upon its performance goals to help determine whether the core strategies are actually achieving its strategic goals. Therefore, NSF focuses on what the four key processes actually produces – intermediate results and not final outcomes.

To evaluate performance of its basic research programs, NSF is developing criteria for what it considers to be a "minimally effective" and a "successful" effort towards attaining intermediate outcomes, with enough precision that actual results can be compared to the predetermined criteria. Its method will be to convene appropriately constituted expert panels to evaluate its performance of its Directorates and determine whether they have been "minimally effective", "fully successful", or perhaps not successful at all.

These panels will use broad aggregates of the Foundation activities, covering all of its research and educational activities in a rotation of panels over a period of several years. The fields examined by those outside panels will span the entire Foundation or several areas assessed in any one year.

NSF intends to provide those panels with the best information available to use in their deliberations. Each panel will make suggestions regarding how NSF can improve performance with respect to the particular goal. Membership on these panels is considered crucial. Members must have sufficient expertise that their judgement carry weight with the educational community. Members must also be independent enough of NSF funding so that their evaluations are considered credible. NSF, therefore, offer the following guidance for selecting members for review:

- U.S. researchers or educators in the field being examined, giving special consideration to those who have not received NSF support.
- Non–U.S. researchers or education in the field being studies.
- Stakeholders, including researchers in neighboring fields and other members whose views represent the prospective by problem–solving organizations, such as state and local government, industrial firms, national laboratories, or non–profit organizations.

The Foundation recognizes a need to collect more information from grantees about the results of their activities and the need to make that information more readily available for performing evaluations. Facilities and centers have standard data bases on their activities.

Facilities also prepare annual progress reports but there is no standard set of questions they must answer in the reports. The Education and Human Resource Directorate has recently established a monitoring database, which promises in depth information about science education activities. But the single largest segment of NSF funding– research projects– currently report only in paper form on open–ended questions, and the forms are scattered throughout the agency including at some offsite.

NSF is proposes to modify the format for these reports:

- Scientific achievements and other outcomes of supported activities including descriptions of perceived significance of the outcomes to the research community and prizes awarded for work performed under the project.
- Human resources of the project, including names and levels of involvement by graduate and undergraduate students.
- Dissemination of results, including publications in respected journals and special efforts aimed at sharing results with stakeholder groups; including description of the perceived significance of the outcome to stakeholders.
- Partnerships involved in or formulated through the award, including international collaborations, collaboration with stakeholder groups, including description of the perceived significance of the outcomes to stakeholders.

The standardized information that collected on the final project report will represent a major data source for developing the Foundation's annual performance report. These would include, for example, metrics such as members of the project involved in collaborations with industrial groups, the range of journals in which publications appear, major prizes and awards received. Independent assessment panels would use the information in aggregate form or it could be used in conjunction with other information to describe performance or to conduct special studies.

National Institute for Standards and Technology

The National Institute of Standards and Technology (NIST), formerly the National Bureau of Standards (NBS), is responsible for developing and maintaining measurements in the U.S. The Organic Act of March 3, 1901, which established the NBS; defined the purpose of the Institute:

> ...the Office of Standards Weights and Measures shall hereafter be known as the National Bureau of Standards...the functions of the bureau shall consist of the custody of the standards; the comparison of the standards used in scientific investigations, engineering, manufacturing, commerce, and educational institutions with the standards adopted or recognized by the Government; the construction, when necessary, of standards, their multiples and subdivisions; the testing and calibration of standard measuring apparatus; the solution of problems which arise in connection with standards; the determination of physical constants and the determination of physical constants and the properties of materials, when such data are of great importance to scientific or manufacturing interests and are not obtained of sufficient accuracy elsewhere.

On July 21, 1950, the Act of July 12, 1894 ("An Act to define and establish the units of electrical measure") was repeated by Public Law 617. Therein was stated:

> It shall be the duty of the Secretary of Commerce to establish the values of the primary electric and photometric units in absolute measure, and the legal values for these units shall be those represented by, or derived from, national reference standards maintained by the Department of Commerce.

Then on July 22, 1950, the Organic Act of 1901 was amended by Public Law 619 to read:

> ...the Secretary of Commerce ... is authorized to undertake the following functions:
> (a) The custody, maintenance, and development of the national standards of measurement, and the provision of means and methods for making measurements consistent with those standards ....

Finally, these responsibilities were transferred to NIST, new name for NBS, under the Omnibus Trade and Competitiveness Act of 1988.

The overall organizational structure at NIST is based on eight laboratories. Within each laboratory are divisions, and within divisions are several research groups. For example, the Electronics and Electrical Engineering Laboratory (EEEL) is one of the eight laboratories at NIST. The Electricity Division is one of five divisions within the Electricity Division.

The National Institute for Standards and Technology (NIST) is reviewed annually by two external review, a general policy and management review, and a detailed technical review. The Visiting Committee on Advanced Technology (Committee) reviews general policy, organization, budget, and programs of NIST. The Committee submits an annual report (NIST, 1991a) which includes reviews of NIST's science, engineering and technology transfer programs.

The Board of Assessment, under the auspices of the National Academy of Sciences (NAS), performs a detailed technical review (NIST, 1996). Seventeen panels of reviewers (about ten person's per panel) from industry and academia conduct program reviews based on site visits at NIST's facilities. The panels address variants of research quality, and because of NIST's unique role in supporting competiveness, pay particular attention to technology transfer, industrial coupling, and emerging technologies while quantitative indicators of research impact are not addressed in the panel's annual report (NIST, 1991b), impacts of the research on technology and competiveness are addressed extensively. Recommendations for improvements in these impacts are provided.

Other Research-Oriented Federal Agencies

In 1989, the DOE performed an assessment of projects funded by its Office of Basic Energy Sciences (DOE, 1982; Logsdon, 1985, Kostoff, 1988). Out of approximately 1200 active projects supported by BES, a randomly selected sample of 129 projects were reviewed by panels of scientific peers. Projects were grouped according to areas of science, and the reviews were conducted on 40 separate days by 40 separate panels, with an average of four members and three projects per panel. The reviewers were, for the most part, bench level scientist independent of the DOE.

The reviewers rated projects according to seven attributes: Team Quality (TQ), Scientific Merit (SM), Scientific Approach (SA), Productivity (P), Importance to Mission (IM), Energy Impact (EI), and Overall Project Quality (OPQ). The three factors which approximate the potential impact of the research were SM, IM, and EI. SM incorporated the potential impact of the research on allied research fields. IM covers the ways in which a research project could contribute to the nation's energy needs. EI was designed to predict the probable impact of the research project on energy development, conservation and use. These attribute scores were assigned weights reflecting the relative importance of each attribute to desired outcomes. The weighted attribute scores were then summed over panel members to determine a total project score. These project scores were then used to rank projects and assess their contributions to the overall mission of Basic Energy Sciences program. After the scoring by the panels was completed, regression models were used to quantify the relationship between the OPQ factor (the summary score) and other rating factors. The regression analysis produce a correlation coefficient of .89, which meant that six factors selected constituted the bulk of the consideration which the reviewers used to score the OPQ rating factor.

The Office of Naval Research's (ONR) review process has a major peer evaluation component adapted to meet the particular needs of the organizational unit under review. The two reviews described here are those used to evaluate ONR's two largest programs, the Research Programs Department (RPD) and the Naval Research Laboratory (NRL).

RPD sponsors basic research is performed mainly by universities, and consist of 13 divisions, organized along scientific disciplines. Each division has two separate groups that contribute to the one day annual review. One group is the Division's Board of Visitors (BOV), whose membership is comprised of academia, industry, and non–ONR government. The majority of the BOV members are from the research community, but typically the BOV members include representatives from the technology or development community and the operational Navy. The other groups contributing to the review are the Research Advisory Board, and the senior management of the RPD whose backgrounds span a wide range of scientific disciplines.

For the review, the Division Director provides an overview of the total Division, including its programs, accomplishments, new opportunities, and management issues. The Division's program managers describe their programs in detail, including the impact on science of their accomplishments, potential or ongoing measures such as publications, and potential impacts on the Navy if the effort is successful. The reviewers complete comment sheets, focusing on scientific merit, technical approach, and potential naval impact, and later discuss their findings with the RPD management.

IV. COST AND BENEFITS OF ENERGY R&D

Classification of Energy R&D Technologies

Energy R&D Technologies are generally classified based upon the ultimate purpose of the investment. The National Science Foundation defines three phases of R&D. The objective of basic research is to gain more complete knowledge or understanding of the subject under study, without specific applications in mind. Applied research is aimed at gaining knowledge or understanding to determine the means by which a specific, reorganized need may be met.

In keeping with the NSF definitions, we classify DOE R&D efforts by phase of the research. Those technological endeavors that are non-specific in terms of application-oriented activities in Basic Research are considered group I technologies. Benefits generated from these endeavors are difficult to quantify directly since those values are generally incorporated into applied and developmental research which produces direct benefits that are more readily traceable. A selected number of technologies that DOE supports through its basic energy research programs is shown in Table 6. To illustrate the type of areas involved in these exploratory research initiatives, a more detailed description for the materials science programs is shown in Table 7.

Perhaps the type of energy R&D activities that appears to generate the most attention consist of those endeavors to develop technologies that have direct end -useobjectives in energy efficiency. These applied energy R&D efforts contribute a major area for DOE's research. A list of some of the major applied technologies or programs is presented in Table 8. A more comprehensive list of applied technologies has been listed in the appendix.

Although most studies suggest that DOE devotes more of its efforts to Applied technologies, this belief depends on how various technologies are classified. Table 6 shows a list of programs DOE considers its Fundamental Scientific Program. While other classifications of energy research limit the agency's basic research to those projects included under the Basic Energy Science Program, DOE's broader list raises the preparation of energy R&D that consist of basic research to approximately equal to its applied endeavors.

Table 7

Classification of Energy R&D Technologies

	Budget Authority (In Millions) FY 1995 Approp.
Fundamental Science Research Programs (Group I)	
High Energy Physics	$642.1
Nuclear Physics	331.5
Basic Energy Sciences	733.9
Advanced Neutron Source	20.8
Biological and Environmental Research	436.6
Fusion Energy	368.4
Multiprogram Energy Laboratories – Facilities Support	43.0
University and Science Education	66.0
Laboratory Technology Transfer	56.9
Analysis and Program Direction	28.5
Technology Partnerships	-------
Subtotal Fundamental Science Research	**2,727.7**
Applied Energy Research Programs (Group II)	
Nuclear Power	1,004.5
Coal	804.3
Other Fossil Energy	95.8
Renewable Energy	243.6
Energy End Use	262.5
Subtotal: Applied Energy Research	**2,410.7**

Source: Budget Highlights

Table 8

Group I Technologies and Fiscal Year 1994 Funding Levels

(Millions of Current Dollars)

DIVISION	# OF TECHNOLOGIES	FUNDING (FY 92)
Material Sciences	445	254.4
Chemical Sciences	437	156.5
Energy Biosciences	259	24.4
Engineering & Geosciences	236	35.4
TOTAL	1,377	$470.7

Source: Tabulation derived from: Material Science Programs (1994) Summaries of FY 1994. Geosciences Research, Annual Report and Summaries of FY 1994 Activities, Division of Energy Biosciences, Annual Report, Division of Chemical Sciences (FY 1994).

Table 9

Group II Technologies And Funding Levels

(Millions of Dollars)

#	TECHNOLOGY	≤1990	1991	1992	1993	1994
1	Control Technology & Coal Preparation	407.9	55.6	50.5	42.6	38.4
2	Advanced Research & Technology Development	603.0	31.7	29.9	26.5	26.5
3	Coal Liquefaction	1,531.3	42.7	39.1	37.4	16.0
4	Combustion Systems	502.1	36.9	37.3	36.7	43.8
5	Heat Engine Program	287.0	23.6	17.9	4.3	0.0
6	Magnetohydrodynamics	686.2	40.0	40.3	30.3	4.8
7	Surface Coal Gasification	935.3	15.3	11.0	10.9	11.9
8	Enhanced Oil Recovery and Advanced Oil Extraction & Process Technology	276.0	41.8	50.7	56.8	80.9
9	Oil Shale	335.0	17.4	5.8	5.6	0.0
10	Natural Gas Resource & Extraction	257.9	15.9	12.4	13.5	17.0
11	Advanced Turbine Systems	0.0	0.0	0.8	9.8	23.7
12	Fuel Cells	487.8	42.9	50.8	51.1	49.3
13	Hydrogen Systems	56.0	15.8	16.0	22.5	27.1
14	Photovoltaics	996.2	46.1	60.0	63.7	78.0
15	Biofuels	391.4	32.9	39.0	47.1	58.2
16	Solar Buildings Technology	627.4	2.0	2.0	2.9	5.0
17	Solar Thermal Energy Systems	882.6	19.1	28.8	26.3	32.7
18	Wind Energy Systems	471.9	11.0	21.3	23.4	30.4
19	Ocean Energy Systems	240.3	2.6	2.0	0.9	0.0
20	Geothermal Energy	982.5	27.1	26.9	22.8	24.0
21	Hydroelectric Power	72.4	1.0	1.0	1.1	1.1
22	Advanced Light Water Reactor	100.8	37.3	61.9	58.7	57.8
23	Advanced Liquid Metal Reactor	958.9	125.7	136.8	136.1	23.6
24	Modular High Temperature Gas Reactor	146.9	18.5	15.8	18.0	-0-
25	Magnetic Fusion Energy	6,137.4	285.0	323.2	323.5	343.6
26	Inertial Fusion Energy	61.8	7.6	9.0	7.7	4.0

Source: Data derived from DOE response to request for J. Bennett Johnson, Committee on Appropriations, June 15, 1993.

Table 10

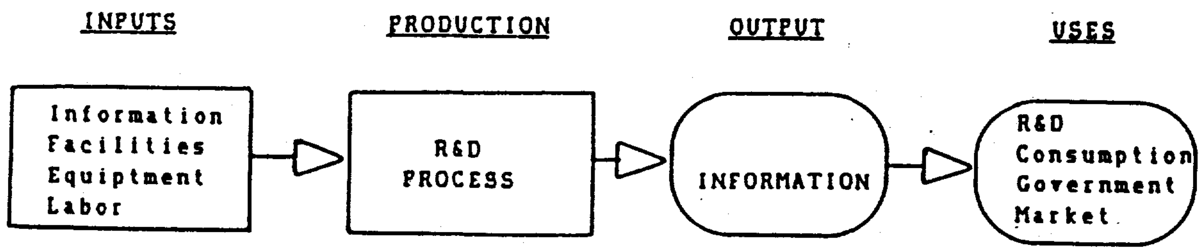
Basic Research Projects Supported by Material Sciences Program (1994)

SUB-PROGRAM	LABORATORIES	UNIVERSITIES	TOTAL
Metallurgy & Ceramics	86	121	207
Structure of Metals	32	45	77
Mechanical Properties	16	27	43
Physical Properties	15	35	50
Radiation Effects	7	4	11
Engineering Materials	16	10	26
Solid State Physics	81	27	158
Neutron Scattering	8	3	11
Experimental Research	52	46	98
Theoretical Research	8	26	34
Partical-Solid Interactions	3	2	5
Engineering Physics	10	-0-	10
Materials Chemistry	32	27	59
Synthesis & Chemical Structure	9	9	18
Polymer & Engineering Chemistry	13	15	28
High Temperature Chemistry	10	3	13
Facility Operations	12	---	12
Small Business Innovative Research	---	---	9
TOTAL	211	175	389

Source: U.S. Department of Energy, Office of Energy Research, Basic Energy Sciences; Division of Material Sciences: Material Sciences Programs, FY 1994 (April 1995).

Exhibit 1 presents a diagram to describe the R&D process. The inputs to R&D include the services from the stock of information relevant to a particular project (knowledge pool), the services provided by the capital facilities and equipment used to conduct the project and finally, all of the different types of labor services provided by those working directly or indirectly on the project. These inputs are combined through a transformation or production function, labeled "R&D" in the center of the diagram, to produce the output of the project, which is information in the case of basic research and some type of technology for applied research. In basic science, this information might take the form of a new hypothesis or the results of a evaluation of an existing hypothesis. In applied research, this information might consist of results from an experimental method for producing an alternative source of energy or by an experiment to determine the combustion properties of a particular material (ceramics). Finally, Exhibit 1 shows that the information produced from the R&D process can be used in one or more of four different ways. First, the information can be reused as an input in other R&D projects. Second, it can be "consumed" by individual scientists for personal enjoyment and professional advancement. Third, it can be used by government to support policy decisions. Fourth, it can be used in private markets by firms to help reduce the costs of existing products or to develop new or improved products for sale in markets.

Exhibit 1. Schematics of the R&D Production Process



Source: Pacific Northwest Laboratory, Assessing the Benefits of Other Research: Three Case Studies. R.S. Nesse, et al., September, 1987.

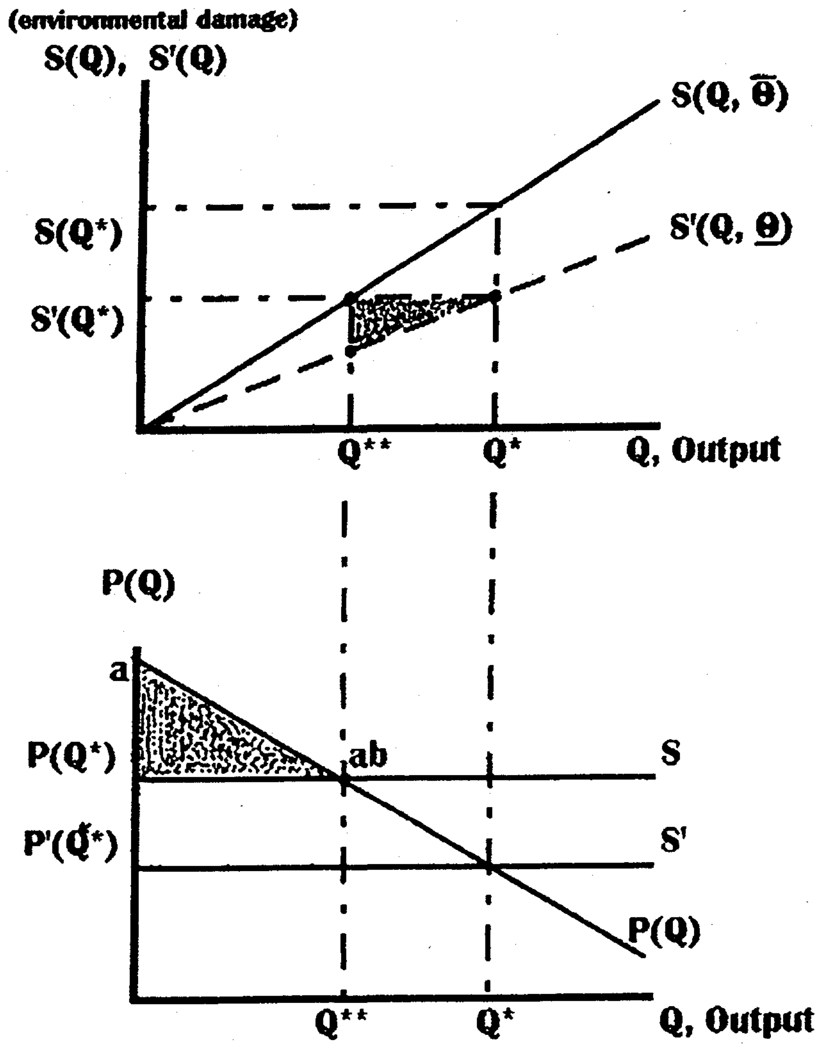
These alternative uses of information represent potential benefits which can, in theory, be quantified. This framework is based on the principle that rational economic agents (i.e., an individual consumer or firm) are willing to sacrifice some amount of resources for an increase in personal welfare. The principle also suggests that there is a minimum – not necessarily the same – amount of resources which a rational economic agent will accept to forego that increase in welfare. The willingness–to–pay principle can be used to measure the net benefits that accrue from research.

Social and Private Benefits

A major portion of DOE's applied research is administered through its Office of Energy Efficiency and consist of research aimed at developing technologies that promises to reduce the amount of energy required to produce a given level of output. Examples include such technologies as low E windows and energy efficient refrigeration/freezer compressors. Figure 1 shows a basic framework through which applied technologies are understood to benefit the nation. The economy's output is assumed to be at Q^*. To produce this level of output, a given amount of fossil energy (coal, natural gas, petroleum) must be consumed. Consumption of fossil energy generates negative externalities in the form of air pollution (SO, NO, CO, particulates). The pollution generated from the production is implicitly represented by function S(Q). Now suppose that DOE is successful in research aimed at developing a new technologies that enables firms to product output Q^* with less energy.

Figure 1

How DOE's Applied R&D Programs Benefit the Nation



This new output–energy requirement function is represented by the function S'(Q). At each level of output, the amount of pollution generated is lower than before the discovery. Thus, the value of DOE's investment, measured as improvements in the quality of the environment, is represented by the shaded area between S(Q) and S'(Q). This value of its technology can also be expressed in terms of output that would otherwise be sacrificed in order to maintain the level of environmental quality. This amount of output is measured along the x axis between Q^{xx} and Q^{x}.

To secure these prospective environmental improvements, however, the value of the added output, $Q^{**} - Q^{*}$, must exceed the cost to firms purchasing, installing, and operating the more energy efficient technology. Since the added output is a flow over time, the increase in sales, improvement in environmental quality and cost must be appropriately discounted to arrive at the net value of the technology. The benefits of the technological development are not limited to firms adopting it. The additional output is sold lower at prices as is depicted in the lower chart. Those already purchasing the output are provided new benefits by being able to purchase the goods at lower market prices. This increases the amount of consumer surplus derived from the product. Also, the additional output enters new utility functions of households that, prior to the introduction of the technology, could not reconcile its price to the additional benefits they derived by purchasing the good. Finally, the value of pollution avoided attributable to use of a more energy–efficient technology must be represented. This value is considered a positive externality and as such it must be added to welfare in order to derive a reflection of the net social benefit to society from the investment.

An energy R&D program also helps to reduce the U.S. dependency on forcing energy sources. This benefits softens the adverse effects that political disruptions in less politically stable oil exporting countries can have on the domestic economy. The possible effects of such situation are so diffused throughout the economy and difficult to identify with any degree of accuracy. Thus avoiding those disruption might be viewed as a public good generated from energy R&D programs.

Finally, a developed energy technology lowers energy requirement in domestic productions. Doing so enhances the competitiveness of domestic firms with foreign rivals. This helps to improve exports and checks the flow of imports. This expands domestic demand for products creating new output, employment and income.

In summary, a developed DOE technology might be viewed as generating several types of benefits: Environmental Implements, Welfare changes in terms of consumer surplus, energy security, and enhanced competition.

One way that the information produced by energy R&D yields benefits is through its effect on the ability of firms to supply goods in markets. This can occur through several paths. The first, and most direct one, is through the effect of applied energy R&D on the production of market goods. A second possible route is through the effect which information has on a government policy decision that, in turn, influences the production of goods and services by firms. For example, more accurate information about the toxicity of a chemical could lead to a decision to allow production of the chemical that had been banned on the basis of the best previous information. Finally, R&D can benefit the market indirectly either through successive phases of R&D which eventually results in the production of information that can be used in the production of market goods and services, or through "spinoffs" to other technologies.

An alternative method of viewing private or social effects of energy R&D is through producer (or, more generally, seller) surplus. As stated previously, producer surplus arises because there is often a difference between the price at which a product can sold in a market and the minimum price acceptable for the good. The change in this measure of net benefits is represented by the change in the difference between the gross receipts of the firm and the firm's total variable cost.[1] Finally, it should be noted that changes in gross receipts, revenues or other measures for the value of products sold by a firm do not constitute a legitimate welfare measure because these measures fail to account for the opportunity cost of the resources used to produce goods. Since these resources could be used to produce other goods, the cost associated with not using them elsewhere (including economic surplus) must be deducted from gross receipts to obtain a legitimate welfare measure.

[1] Economists often use the term quasi-rent, which measures the difference between the amount which the factors of a resource owner earn in their current occupation and the minimum sum he or she is willing to accept to keep them there, is often used interchangeably with the term producer surplus. Under most conditions the two measures are equivalent.

Figure 2

Net Benefits to Firms of a New Technology

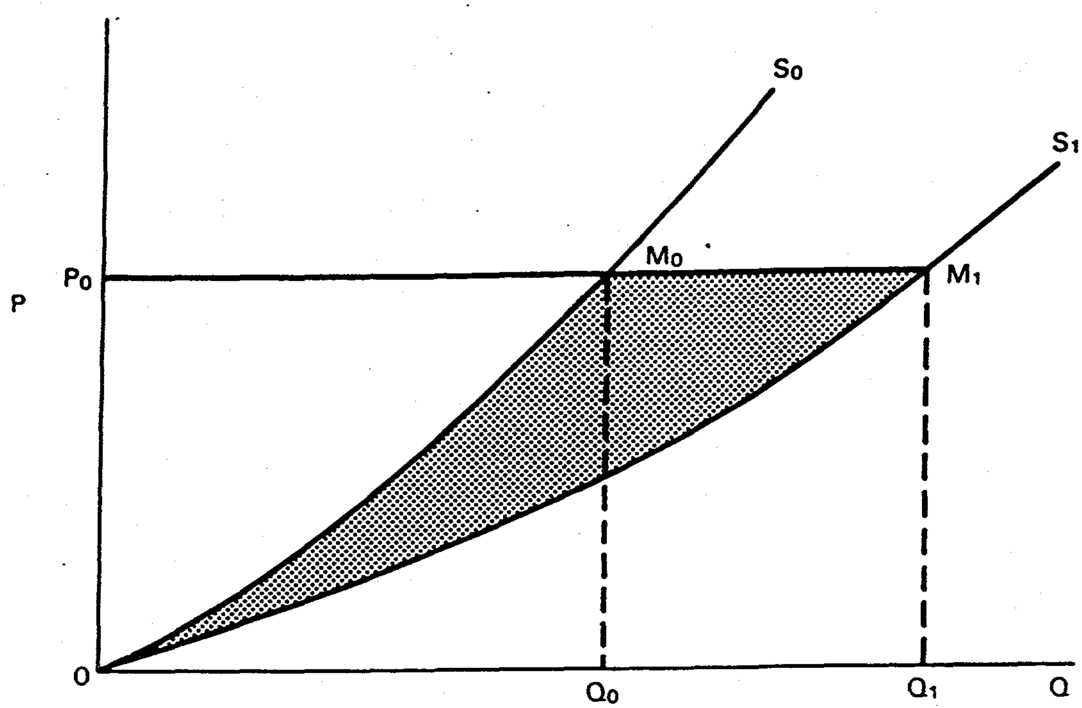


Figure 2 depicts how the change in producer surplus of firms can be used to measure the effects of a government investment in R&D resulting in production of a new technology that makes it possible to reduce the cost of providing a market good or service. The example used involves a actual R&D investment which results in a new technology for reducing the amount of energy used in production of some good **Q**. Let us assume here for simplicity that the relevant market good affected by the R&D investment is additional cubic feet of natural gas produced from drilling by invention of a new techniques for oil and gas exploration. The initial market supply for natural gas is represented by the line OS_0. This curve shows cubic feet of natural gas (Q) on the horizontal axis that firms are willing to offer at corresponding prices (P) on the vertical axis. For simplicity, the market demand for natural gas is assumed to be perfectly price elastic: a change in the quantity of natural gas supplied will have no influence on the market price for each cubic foot of natural gas, P_0. Under these conditions, there is a unique market equilibrium at point M_0, such that firms sell and buyers take Q_0 cubic feet of natural gas at a price of P_0. Gross receipts to sellers of natural gas is P_0 times Q_0. This is represented by the area $OP_0M_0Q_0$. The total variable cost of all firms in the market is measured by the area under the supply curve up to M_0, which is OM_0Q_0. Producer surplus is measured by the difference between gross receipts and total variable costs, which is equal to the pie–shaped area OP_0M_0.

Now suppose that the effect of development and installation of a new technology (mud pulse telemetry) is to reduce the variable cost of harvesting the gas, no matter how many cubic feet are given. The effect of this investment on the availability of gas in the market is shown by the new market supply curve for this product, represented by the line OS_1. This supply curve shows that firms are now willing to provide more natural gas at the same price as before the technology was applied. The result is that, given the same perfectly elastic aggregate demand for natural gas, firms now sell and buyers take Q_1 cubic feet from the market at a price of P_0 cubic foot. Following our previous calculations, producer surplus can now be measured by the area OP_0M_1. The net benefits of the energy R&D investment to firms is measured by the shaded area in Figure 2, OM_0M_1. Conceptually, the change in producer surplus shown in Figure 1 is composed of two parts. The shaded area to the left of Q_0 represents economic surplus due to the reduction in the variable cost of producing natural gas, holding the number of cubic feet harvested at the initial level. The shaded area to the right of Q_0 represents the economic surplus generated by the provision of new support (i.e., $Q_1 - Q_0$), using the new technology.

A second avenue through which a new energy technology produces benefits is through its impact on the welfare of consumers. This can occur in at least three different ways. First, consumer welfare will generally be improved if the energy R&D Program results in new or improved products or in a decrease in the price of existing goods or services. These welfare gains can occur as a result of private or public use of new energy technologies developed through public support of energy R&D (See Exhibit 1). For example, the decision to utilize a specific energy source on the basis of new information discovered through basic energy research might lower the costs of products and services using this newly developed source. Nor is this form of welfare gain limited to goods and services provided in markets. For example, R&D investments that leads to information that results in less expensive pollution control technologies can improve the welfare of individuals in the form of greater enjoyment from a cleaner environment. Second, the welfare of individual scientists can be improved through direct monetary compensation in the form of higher pay, additional grants, and indirectly through greater status in the profession. Finally, the welfare of individual scientists can also be improved through non–monetary compensation as a result of the enjoyment derived from research. This last welfare gain can be extended to non–scientists, as well. By definition, any enjoyment derived from learning about the information provided by R&D represents benefits.

Measuring these types of welfare gains in dollar terms is more complicated for individuals than for firms. One possible measure is ordinary consumer (buyer) surplus, which is defined as the difference between the maximum amount a consumer is willing to pay for a good, rather than to go without it, and the amount actually paid for the good. The primary advantage of this welfare measure is that it can be estimated using information obtained directly from the ordinary demand curves of consumers for goods believed to have benefited from energy R&D. The chief disadvantage of this measure is that it may not always provide a unique measure of the net benefits to individuals associated with R&D investments. Fortunately, Willig (1976) has shown how information obtained from ordinary demand curves can be used in conjunction with consumer surplus to minimize the errors associated with the non-uniqueness problem.

Figure 3

Net Benefits to Individuals of a Price Change

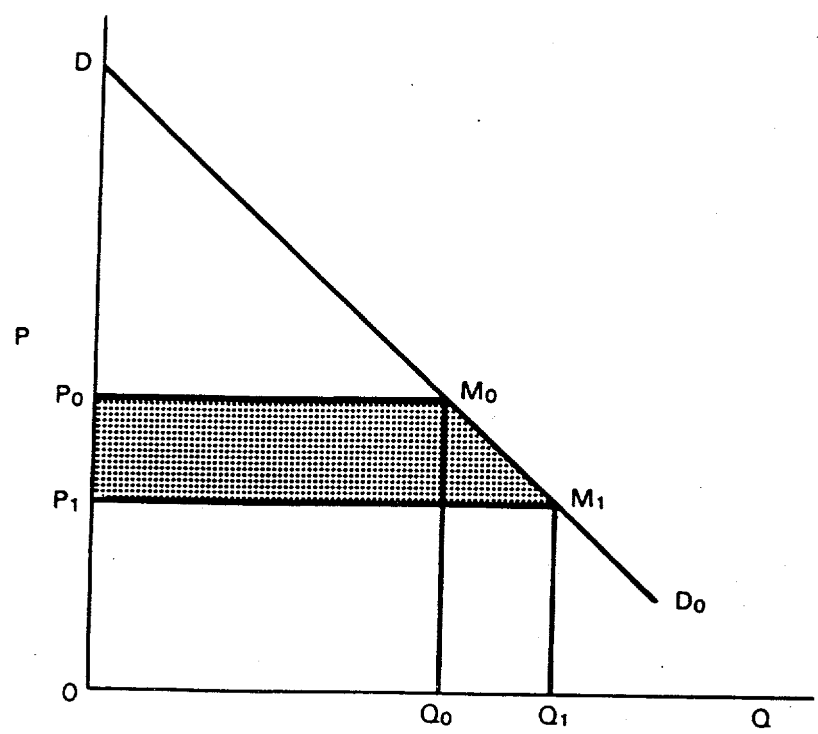


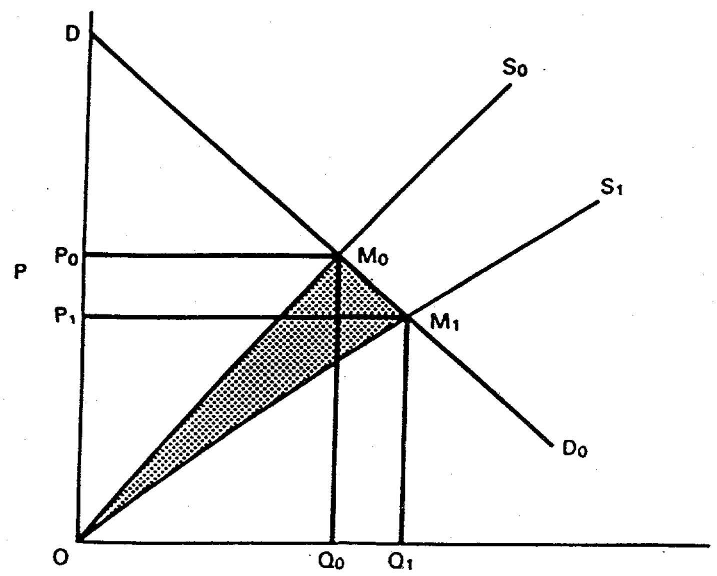
Figure 3 illustrates how the concept of ordinary consumer surplus can be used to quantify the net benefits to individuals from an R&D investment which reduces the price of a market good from P_0 to P_1. It is assumed, for simplicity, that market supply for the relevant good is perfectly elastic. Prior to the application of the information produced by an R&D investment, this supply curve is shown by the horizontal line P_0M_0. The ordinary market demand function for the good is shown by the line DD_0. Each point on this curve describes the quantity of the good (Q) that buyers will take from the market when faced with a specific price (P). Under these conditions, consumers take Q_0 units of the good from the market at a price of P_0 per unit. The consumer surplus associated with this market equilibrium can be calculated as follows. Some individuals are willing to pay as much as D for the first unit of Q rather than do without it. However, they only have to pay P_0, not D, to purchase it. Therefore, the consumer surplus associated with the first unit of Q is equal to $D - P_0$. Repeating this calculation for each additional unit of Q up to Q_0 gives a consumer surplus total which can be represented by the area under the demand curve and above the market price line at P_0. This area is equal to the area of the right triangle bounded by the points P_0DM_0.

Now, suppose that information produced by R&D results in a downward shift in the aggregate supply curve for Q from P_0M_0 to P_1M_1. Under these conditions, Q_1 units are bought and sold in the market at the new price P_1. As a result of these changes consumer surplus increases by an amount represented by the shaded rectangular area, $P_1P_0M_0M_1$, in Figure 3.

The concepts of producer surplus and ordinary consumer surplus are brought together explicitly in Figure 4. This figure shows ordinary demand curves for a market or non-market good can provide information necessary to approximate the net benefits of an R&D investment to both individuals and firms. The initial market supply and demand curves for the good (Q) are shown by the lines DD_0 and OS_0. Market equilibrium occurs at the intersection of the supply and demand curves at M_0 and is characterized by the market clearing price and quantity combination of P_0, Q_0. At this point, ordinary consumer surplus is equal to the area P_0DM_0 and producer surplus is equal to the area OP_0M_0. The effect of new technology on the production of Q is represented by a shift in the aggregate supply curve from OS_0 to OS_1, indicating that the unit cost of producing the good is lower at all levels of output. As a result of this new technology, a new market equilibrium is reached at point M_1. At this new equilibrium, the market price has fallen to P_1, while output and consumption have increased to Q_1. Consumer surplus is now represented by the area P_1DM_1, while producer surplus is represented by the area OP_1M_1. The change in producer and consumer surplus as a result of the new technology is equal to the shaded area OM_0M_1.

Figure 4

Net Benefits of Government Investment in R&D



These benefits that are estimated using consumer surplus represents an approximation of the net benefits received by individuals and firms as a result of the new information in a single market in a single period. However, the results of R&D have the potential to influence the welfare of firms and individuals in a variety of market and non-market situations. Changes in the sum of producer and consumer surplus in a single market is an appropriate measures of only the direct effects of new information developed from a R&D program on economic well-being in a partial equilibrium framework. These same measures are also appropriate in a general equilibrium framework in which markets are linked by the exchange of inputs and outputs, and the effects of new information in one market can spill over into others in the form of changes in price, output and consumption levels. In this more general context, however, aggregation of consumer and producer surpluses must be undertaken with care. This is because an economic agent can be a buyer in one market and a seller in another. Consequently, the impact of a technology is the sum of the buyer and seller surpluses in all markets will generally differ from the change in the sum of the surpluses of all buyers and sellers (Hueth, et al. 1982). To avoid double counting of surpluses, the second method of aggregation is correct.

Social and Private Costs

There are basically three kinds of costs associated with R&D investments and the effects of these investments on the welfare of firms and individuals. First, there are the direct costs associated with the use of new information and other inputs to produce goods and services. The treatment of these types of costs is discussed in conjunction with the measurement of producer surplus since they are reflected in the supply curve. The above discussion of consumer and producer surplus is also appropriate for a second type of cost – costs that may be represented as negative benefits. For example, DOE sponsored R&D may produce information that the energy savings associated with a particular technology were more significant in terms of environment quality than originally believed. If production was low based on the use of the newly developed energy technology, this would have the effect of shifting the supply curve in Figure 3 to the right– for example from S_1 to S_0 – causing an increase in the sum of producer and consumer surplus in the market for that good or services that relies on the technology. On the other hand, if the decision to adopt the energy technology was made strictly on the basis of a narrow concepts benefit–cost analysis, this presumes that the increase in consumer surplus in the market for the technologies might be offset by decreases in producer surplus due to higher cost of adopting the technology.

Finally, there is a need to recognize development costs for an R&D project. The appropriate way to treat this category of cost is to capitalize it. It is the opportunity cost of the resources used in the process of producing information that ultimately leads to development of a technology that is commercially applied. The opportunity cost of these resources measures is that which society must sacrifice in order to fund an R&D project rather than use project resources in their next–best alternative. These costs include all of the costs normally associated with federally funded R&D efforts. For example, they includes the amount DOE is invoiced for the research services provided by the grantee or vendor. They also include costs that are not normally accounted for as R&D costs such as the value of the time spent by DOE officials to screen, evaluate and monitor an R&D program. Opportunity costs also include costs associated with the use of goods and services that may be provided "free" to a project. For example, some laboratories provide materials and chemicals for experimental use by researchers in other labs at no charge to users. However, these goods are valuable to society, even if their only use is experimental, since the resources used to produce them could have been used elsewhere in society. Accordingly, free goods and services should be priced at their best alternative use, which in most cases is not zero.

In the above sections we discussed the importance of a comprehensive measure of benefits and costs associated with the production of new information and development of new technologies. In addition, we tried to suggest that this framework was broad enough to measure changes in benefits and costs associated with a wide variety of uses to which new information and technologies could be put. This includes abstract measures of benefits associated with the enjoyment of science by scientists and others, with changes in the risks to which individuals are exposed, and with changes in environmental quality. Valuation of these so-called non-market activities is controversial both for methodological and normative reasons. In this section, we examine the more important and very real methodological problems associated with valuing these benefits in an applied framework. These problems can be grouped under two headings: (1) those associated with attributing the benefits and costs of R&D, and (2) those related to estimating the benefits and costs of R&D. Normative issues of whether it is right or wrong to convert all values into monetary units is beyond the scope of this report.

Evaluating performance programs, such as energy research involves conducting a hypothetical experiment to determine net welfare gains to society with and without a specific group of research projects. As such, one of the first steps consists of identifying all of the potential effects of the R&D investment program, both favorable and unfavorable. It also involves identifying the market and non-market contexts in which these potential effects could occur and the economic agents (i.e., firms, consumers, factor owners) who will be influenced directly or indirectly in these markets. To see why attribution of benefits is a serious problem in the evaluation of energy R&D, let us first examine case in which the problems are not as severe as in the case of applied R&D and then compare it with basic R&D.

Consider, again, program evaluation to support a decision to finance development of a technology that could lower the cost of drilling for natural gas. Output is a specific technology that avoids the cost of withdrawing a drill to change direction. The immediate beneficiaries of the technology is the oil extracting industry, who will experience increases in producer surplus due to higher yields, lower variable costs, and harvest of more profitable gas. If this technology results in lower market prices for natural gas, then consumers would also benefit through increases in consumer surplus. However, some affects of developing the technology may be unfavorable. Lower market prices for natural gas owing to increased productivity could hurt coal producers by encouraging substitution and cause them to experience a decrease in producer surplus. In addition, the technology could adversely affect wildlife habitats, scenic values, and existing forms of recreation is the national gas region. These negative consequences of the technology would be accompanied by decreases in the consumer surplus of individuals whose use of the environment would be impaired by the project. Finally, the project could also reduce the consumer surplus of individuals who feel unhappy about the environmental effects of the project even though they may not experience them directly through their use of the environment. Attribution of the benefits and costs (i.e., negative benefits) in the above case is relatively straightforward, with the possible exception of the final category of negative benefits. In almost all instances, we can identify the potential (not the actual) consequences of the project and relate these consequences to specific groups of economic agents. Furthermore, this is true whether the evaluation is conducted on an ex ante or ex post basis.

Now consider a near–polar case involving a basic science research project in purely theoretical fields, such as the so–called "unified theory," which, among other things, attempts to trace all of the currently known physical forces back to a single force that was present at the moment the universe was formed. The major problem with evaluating the most basic types of research is that the information produced by such endeavors has no clear effect on any currently available technology. As such, there is simply no way to attribute market–related benefits to such a project without the benefit of a hundreds of years of hindsight. A more immediate effect of such a project will be to increase the stock of knowledge available to other theoretical physicists. This would increase the consumer surplus of individuals, presumably scientists who enjoyed reading or knowing about the results of the project. Identifying the users of information which has not yet been produced may be a somewhat arbitrary exercise. Finally, if basic research is successful, it could indirectly lead to additional monetary compensation and professional recognition for the project team members. While these types of benefits are easier to attribute to individuals, most economists and scientists would be understandably uncomfortable with the use of such a narrow measure as the sole basis for evaluating basic research. Part of the problem lies in the ex ante nature of the evaluation. While the problem of attribution is less cumbersome when applied on an ex post basis, it by no means disappears. In the case of basic research, the results of a research project may be a proof of a mathematical theorem whose only foreseeable use is as an input to other, equally abstract theorems.

The problem of attribution also arises in the context of more applied forms of research. Consider, for example, problems associated with attributing benefits and costs of R&D investments in nuclear medicine. In an ex ante evaluation framework, one encounters problems of attribution similar to those associated with basic research: identifying potential market and non–market benefits. Even market benefits of the most applied forms of R&D are difficult to predict in advance. This is because the link between R&D and its eventual commercial application frequently depends upon advances in other, seemingly peripheral, technologies. The same problem exists, to a degree, in ex post evaluations of applied R&D due to yet to be commercialized applications. Perhaps more serious than this is the effect which different assumptions about the time when an R&D project began can have on the attribution of benefits and costs. For example, if one is attempting to calculate the benefits and costs of R&D in nuclear medicine in an ex post framework, must one include the cost of the Manhattan Project? Presumably, the results of such an investigation would be extremely sensitive to any such assumption.

The problems noted above can generally be traced to one of three sources. First, it is frequently very difficult to fully appropriate the benefits from R&D in private markets. This is because ideas have an illusive quality which causes problems for the "owners" of these ideas to exclude other individuals even through legal protection from using the idea or information. Difficulties in establishing and enforcing ownership rights to information translate into problems with attributing the benefits associated with that information to identifiable sources. Second, information, once it has been produced, takes the form of a "public good" in the sense that its availability to any member of society does not preclude reduce the amount that could be made available to others. Consequently, the benefits of this information can be shared widely and equally by many different economic agents without any real way of tracking all of the benefits resulting from a specific research program. Finally, it is often the case in the public sector that markets do not exist for trading and valuing the information produced by R&D. This further obscures the path of R&D from a research project to its many different uses.

Several fairly general conclusions can be drawn from our discussion of the problems associated with evaluating energy R&D programs. The first is that, except in cases where the primary effects of the R&D consist of direct and identifiable market goods and services, ex ante evaluations are not very productive. This is due to the problems associated with predicting and tracing the effects of R&D from a specific project to all of its potential beneficiaries.

Second, evaluating R&D in fundamental research is likely to be of limited use in assisting decision makers to allocate resources either in an ex ante or ex post framework. This is due, in part, to the fact that results of much fundamental basic science may have limited direct applications, with the exception of the benefits it produces for scientists. The importance of these types of benefits cannot be understated, however. They can be evaluated through less conventional peer review methods. This approach seems likely to be most useful in valuing research that is difficult to trace to direct applications but does reduce the cost of conducting future research.

Third, the fact that the major effects of an R&D project may not be measurable by market values is not a valid *a priori* reason for dismissing it as useful. Recent methodological developments in the field of non–market valuation make it possible to evaluate these effects. The important requirements that must be met are that the primary non–market effects can be identified and traced from the project to specific groups. The availability of data will then determine whether these non–market effects can be valued directly through observed changes in labor compensation or property values, or indirectly through the use of survey methods to elicit individual willingness–to–pay responses.

Finally, if the effects of R&D are to be evaluated on an ex post basis using monetary values (market or non–market); it is important for to place these effects in an appropriate context so that users of this information can provide a better understanding of how important these effects are in relation to others, which, for whatever reasons, have not been quantified. In addition, it is also important to clearly state the assumptions required to conduct the analysis, how sensitive the results of the analysis may be to changes in the assumptions, and whether these assumptions provide an upper or lower bound on the net benefits associated with the effects of R&D that have been measured.

V. METHODOLOGY FOR EVALUATING ENERGY R&D

This section describes methods available for evaluating R&D programs and provides examples of the use of these techniques within research–oriented federal government. The techniques: 1) peer review; 2) bibliometric techniques, 3) citation analysis, and 4) human capital are methods suggested for evaluating Group I energy R&D programs. Applied energy programs may be evaluated by relying on case study, consumer surplus, production function, accounting or user evaluation methods.

Evaluation Models for Group I Technologies

Evaluation methodologies for Group I technology programs (basic research) requires a flexible set of indicators on short–term program outputs that are significant only because of their connection to longer–term outcomes. For constructing summary indicators, all evaluation methods have both strengths and limitations. Many of the indicators of research program outputs could find useful applications in the context of a full–blown program evaluation, but have more severe limitations for use as summary performance indicators. For example, a full–blown evaluation can take into account descriptive analysis of interview data, complex models of program operations, or sophisticated citation analysis. All of these can provide performance–related information to inform an evaluation report, but do not match GPRA's requirements for simple performance indicators.

The purpose of DOE's fundamental science research is to increase understanding of a physical, social, or technological phenomenon related to energy. While understanding itself is hard to measure, knowledge production has proven to be at least in part quantifiable. Two aspects of the knowledge produced through research should be of interest to program managers:

quantity of knowledge produced through the program, quality of that knowledge, and relevance of efforts to goals. This section offers some possible methods to gauge quantity, quality, and relevance of outputs from basic research programs.

The Keystone Model

What model can be used to conceptualize the process by which energy R&D operate to produce benefits that are of value to the nation in a way that permits tradeoffs among programs? Cozzens (1994) suggest an apparatus for articulating outcomes from the research process which she calls the keystone model. The basic premises underlying this model is that research produces intermediate results or unfinished outputs. The value of those outputs are not easily recognizable by people such as members of the Congress. The contribution of basic research to the national agenda can be viewed similar to the influence of investment to growth of the economy. While capital goods are themselves produced through the production process, those outputs are not recognized by the general public as satisfying final demand. Instead, the goods that are produced by the future use of the capital stock is what the general public recognizes as benefits. The services of capital is embodied in a new and more efficient stock of capital goods.

In the keystone model, the knowledge base is considered as the capital stock. New discoveries adds to the knowledge base and enhance its usefulness for applications – oriented activities. Thus basic research influences social welfare not directly but instead indirectly by maintaining the quality of the knowledge base.

The basic schematics of the keystone model is presented in Exhibit 2. A keystone is placed at the top of an arch. If it is removed, the arch falls down. Using this analogy in the context of program evaluation amounts to asking: If a particular R&D program were removed for the set of programs supported, what arches would fall down? What processes would no longer be undertaken? What benefits would no longer be provided? In Exhibit 2, arches represent three types of benefits that are supported by investments in basic research programs. Not all programs support all three types of benefits, but most programs support more than one. One arch is the knowledge base, another is the application of that knowledge base to problem-solving, and a third is education - which is generally transmitted to the economy through students practicioners:

Exhibit 2

Cozzens, S.E., "Strategic Evaluation and the Keystone Model of Basic Research," (April, 1994).

There are two benefit flows in the model: a knowledge and people flow. The primary function of basic research programs is to enhance the quality of knowledge available through the research front in a particular field. Research front knowledge goes into larger pools drawn upon by practitioners and educators. Only in special instances are the research front knowledge used directly in the economy. The quality of information contained in the knowledge pool is determined by whether it is used regularly by mission–oriented research activities.

The effectiveness of a basic research program is not well represented by output measures since R&D and technical progress are, in themselves, rarely ends. R&D is usually conducted because it contributes to higher goals, whether those goals are company profits, the quality of the intellectual capital stock, environmental quality or leadership in science: DOE typically want to know whether a particular program has contributed to these higher goals. As a consequence, R&D evaluations are conducted in order to link R&D technologies to impacts on some other social, economic, or technical measures.

This is the most challenging aspect of R&D program evaluation because it is difficult to capture outcomes and impacts accurately. This methodological issue is complicated by an additional factor, namely the difficulty of isolating the observed effect of a particular R&D program. Impact assessment not only evaluate the effectiveness of R&D, but the overall innovation and technology delivery system (Hill and Hansen, 1988; Brown and Svenson, 1988; Collier, 1977; EZRA, 1975).

R&D impact analysis is of a uniform sort. In almost all instances, it involves either estimates of return–on–research or R&D cost–benefit ratios, although a couple of methods described by Schainblatt (1982) include of algorithms that lead to project rankings. Returns/benefits can be either actual or expected, but estimating expected benefits requires projections (rather than actual) for market value or payoff. Almost all techniques additionally require estimates of operational costs (R&D, start–up, and/or production). Two other methods, explained by Brown and Svenson (1988) and Patterson (1983) are more refined in that they identify categories of potential benefits and estimate economic benefit ratios.

A description on various methods for evaluating R&D programs is provided in this section. Generally, formulas for calculating the potential return from an entire R&D program are discussed by Collier (1977) and Collier and Gee (1973), whereas Patterson (1983) outlines methods for estimating rates of return to those technologies that are, in part, supported by public funds. Porter (1978) provides a similar method for computing a "research–return ratio", but rather than using constant estimates for variables such as market sales, costs, etc., an estimate of the probability of market success for the innovation is incorporated.

The work of Grupp, Schmoch, and Kuntze (1991), represent probably the only technologies competitiveness evaluation conducted for the purpose of resource allocation. Moreover, it was conducted for government–performed R&D programs, a rarity in the evaluation literature. The authors discuss a method where patentable outcomes of R&D are linked to a broader technological space using patents "symbols" (the European equivalent of patent classes) and keyword mapping. Once the relevant population of patents are defined, a variety of analyses can be conducted on different elements of patent data. The analysis allows the theory to determine the technical results of an R&D program (quantity, quality, direction) compared with other nations. They then illustrated how the results could be used to allocate R&D funds, dependency upon whether the government wanted to boost different programs or maintain the strength of vita ones.

Because of uncertainties attached to each and every evaluation procedure, the evaluation of basic research is best done by using ex post review and citation evidence jointly. The steps in such an evaluation process are as follows:

1. For a given portfolio, take a sufficiently large random sample of completed projects. The sample may be stratified so that information about sub-fields can be constructed (Logsdon and Rubin, 1985; Kustoff, 1988).

2. Design a scoring system to be applied to projects. Scoring systems can be aggregated or disagregated. In an aggregate scoring system, evaluators make an overall summary judgement about completed projects the way they do, for example, in the NSF ex ante peer review system. Reviewers assess the quality of a project on a point scale ranging from excellent to poor.

3. Compare the overall ex-post peer review of particular projects with available bibliometric information.

In making these comparisons, adjustments may be necessary to allow for variations in project vintage since a more recent projects may not have produced many publications and citations, despite obvious technical quality, and the results from a vintaged projects may have been so assimulated that users no longer cite it. In additions, adjustments may have to be made for critical or negative citations some science currently judged as not meritorious will turn out to be so and vice versa, and so evaluations should be periodically revisited.

Peer Review

A review of methods used in selected federal agencies to evaluate basic research investments indicates that most agencies base research funding decisions on peer review rather than on any economic or other quantitative method (Logsdon and Rubin 1985). The agency first to experiment the peer review method for funding basic research was the Office of Naval Research (ONR), a research agency within the Department of Defense. The ONR practice was based on an earlier recommendation by Vannevar Bush (1945) that peer review by independent scientists, having no direct dependency on the federal government, would strengthen basic research by separating the research mission of federal agencies from their operational missions. The ONR peer review model constituted a multi-level process, involving both internal functional and external peer reviews. This model provides the basis for current-day peer review procedures used by the National Science Foundation (NSF), the Office of Energy Research in the U.S. Department of Energy, and other federal agencies.

Other federal agencies that rely the peer review process to evaluate basic science funding decisions include the National Institute of Health (NIH), NASA's Office of Aeronautical and Space Technology and the Army Research Laboratory (ARL). As a recent review of research evaluation methods by the U.S. Office of Technology Assessment (OTA) suggest, there is a high level of confidence in the peer review process both in these agencies and in the scientific community (OTA 1986).

The peer review process can be used as a standard for allocating funds across research projects, for exercising managerial control over an existing portfolio of projects and for making decisions regarding continuation of funding. Peer review has also been the traditional mechanism through which agencies have used to justify their research to various oversight groups. Most agencies use the traditional form of peer review in which outside scientists are asked to assess various attributes of a proposal, project, or program using qualitative measures of performance. However, there have been efforts to make the peer review process more quantitative by using numbers.

These attribute scores are assigned weights reflecting the relative importance of each attribute to desired outcomes. The weighted attribute scores by panel numbers were then summed to develop a total project score. These project scores were then used to rank projects according to their contributions to the overall mission of Basic Energy Science program.

The peer review model has a number of advantages, perhaps most important of which is that it has broad support. Both those who administer the process and members of the scientific community, whose own research is often subjected to peer review, consider it to be an efficient process for making basic research funding decisions.

According to the OTA study, there is far less agreement regarding the validity of using economic or bibliometric methods as a basis for deciding what research to fund. This lack of agreement is explained in part by the fact that scientists within the same discipline participate in the peer review process both as reviewers and as perspective beneficiaries , whereas other forms of evaluation are more likely to be conducted by professionals outside the discipline of the proposer.

A second aspect of the peer review method that makes it valuable is that it helps to give research programs a sense of scientific credibility they might otherwise lack if research funding decisions were made without the advice and consent, so to speak, of the scientific community. The importance of gathering a consensus in funding decisions helps to explain the broad support for peer review within the scientific community. As mentioned above, most federal agencies that employ this method do not use highly quantitative peer review methods by assigning scores or to rank research proposals, projects or programs. Rather, these agencies rely heavily upon the weight of consensus among multiple reviewers with the outcome decision as either to approve or deny. Thus, approval of a technical proposal or project through the peer review process generally signifies broad agreement within relevant disciplines.

Finally, peer review is an extremely flexible approach to evaluation when the outputs of a research proposal or project are highly abstract and not immediately amenable to a commercial application. As such, it is best employed in evaluation of the type of energy research technologies under the agency's Basic Energy Science. The application of this approach to basic research can be defended on the grounds that it yields decisions that are presumably consistent with the preferences of those who tend to make the most immediate use and derive the most immediate satisfaction from the results of the research. While no rigorous efforts have been undertaken to determine the value of information for its own sake, discrepancies between academic and industry salaries in many disciplines certainly lends support to the view that many scientists are willing to sacrifice substantial amounts of money (i.e., foregone income) for the satisfaction afforded by intellectual pursuits. By contrast, the expected value of the research in its future commercial application is likely to be very small due to a combination of uncertainty about future uses and values and the impact of discounting benefits achieved in the very distant future.

Martin and Irvine's has been applied the peer review method to evaluate performance by laboratories engaged in high energy physics (1984) and radio astronomy (1983b). Although, their method is relatively controversial and has not been used by any agencies of the federal government to evaluate basic research, more general bibliometric studies have been used by federal agencies.

The peer review methods is often criticized on four grounds. The most common criticism directed at this approach is that unless objectives of programs are all linked to a common good, it cannot be used to compare the value of research across R&D programs for the purpose of resource allocation. In fact, most peer reviews do not offer a single–valued quality metric for allocating scarce research resources to competing research programs. This can lead not only to ambiguity about how limited resources are to be allocated among a number of technically exceptional programs, but also to concerns about the nature of the criteria used to make these incremental decisions.

A second concern is that the results of peer reviews tend to reflect the preferences of only the individual peer reviewers, along with their backgrounds, biases and objectives. Whereas the results of the research might benefit a much broader group of people whose preferences are not taken into account.

A third criticism often leveled at the peer review method is that it is risk averse in that it tends to promote what Thomas Kuhn (1962) terms "problem solving" in science rather than invention. This manifests itself as a tendency among peer reviewers to favor research methods that are well–accepted over more controversial approaches.

Finally, peer review is often criticized because it may be more subject to manipulation by agency administrators who may have a particular interest towards particular research result that they want to achieve. These personal biases may be easier to achieve through peer review than through other forms of evaluation because of the discretion offered in selecting the reviewers and assigning weights to different attributes of a program. However, while these and other questions about peer review persist, no major proposals for change have been convincing enough to warrant a substantial overhaul the peer review process.

Bibliometric Methods

As pressure for quantitative measures of the performance, impacts, and outcomes of scientific efforts increase, the demand for quantitative yardsticks for output of science has become more urgent even when it is apparent that the basic measure of output is qualitative. This section outlined the methodology, bibliometrics, the use of publication–based data to evaluate energy R&D. The application of econometric analysis to bibliographical data is the design of the method.

Bibliometrics represents study and analysis of scientific output as evidenced through publication–based data. It is a clearly delineated body of research involving the measurement of "physical" units of publications, bibliographic citations, and derivatives of them (Broadus, 1987). Pritchard defined the methodology as the application of statistical methods to books and other methods of communication to quantify output of research" (Pritchard, 1969). In practice, this means that the number of publications or citations attributable to a research program may be used to judge the productivity, or output, of the research program and by aggregation across programs, an entire organization. Thus, bibliometrics essentially serves as an approximation of the output of R&D primarily due to the immense difficulties and uncertainties with direct evaluation of research and development activities.

Although attention to output of scholarly activities may be traced to the turn of the century, formal bibliometric analysis is said to have originated in the 1960s. Derek de Salla Price and Eugene Garfield pioneered the movement to develop bibliometric indicators (price, 1963; Garfield, et al., 1964). Henry Small helped to refine the method with the development of co-citation analysis (Small, 1974). They and others recognized the need to quantify the output of science, so that those outputs could be both predicted and monitored (OTA, 1986). This development represented a departure from the more subjective methods in use at the time. Bibliometric provided an alternative body of information about the scientific community that was not uniformly available through methods such as peer review and descriptive accounts of scientific developments.

The creation of the Science Citation Index (SCI) in 1961 enabled bibliometric analysis to become even more systematized (Garfield, 1979). Prior to the development of the SCI, publication counts and citation analysis were highly labor-incentive. Computerized analysis allowed for larger, faster, and more complete analysis of publications and citations. The SCI has subsequently become the central data source for applying quantitative analysis to science (OTA, 1986). The database provides information in three main units: publications, citations, and authorship. Furthermore, each journal can be classified according to disciplines and specifications (Narin, 1976). These units of analysis can be used to accomplish several different analytical ends.

Bibliometric methods consist of those methodologies that attempt to measure the quantity and quality of the output of a research project, program or institution by counting publications, citations or cross-citations associated with technical projects. The important assumption that underlies this approach is that the output of research is information and that the contribution of a supported technology to the knowledge base of society can be measured by the number and quality of publications derived from it. A variant to this approach combines bibliometric methods of peer review in an attempt to assess the efficiency of research investments, i.e., number of publications and citations per dollar of investment. Bibliometric methods, like peer review methods, have been criticized because the expost nature it is difficult to use them to make resource allocation decisions involving comparisons between research and other federal programs.

Early bibliometric efforts explored the feasibility of understanding science through its literature, independent of the scientists themselves. The first bibliometricians tended to use counts of citations as an indication of the directions in which science was moving. However, according to Chubin (1976, 1981), limitations associated with measuring and scaling these outputs soon went beyond simply counting of citations to more complicated statistical and mathematical techniques that would allow bibliometricians to describe, in quantitative terms, the structure of the information base reflected in the scientific literature. Now that these tools have been developed, bibliometricians are attempting to use them to evaluate research projects and programs on an ex post basis.

Although "bibliometric analysis" is often used generically, it is represented by several different technical approaches, each of which have different purposes and utilities. The common denominator among there forms is the concept that publications represent a flow of information primarily from basic science (Martin and Irvone, 83). Each of these bibliometric approaches provide different information on the scientific enterprise. The main derivatives of bibliometric are: publication counts, citations counts, co–citation analysis, co–word analysis, scientific "mapping", and citations in patents.

Publication Counts: This bibliometric approach involves counting scientific publications published by researchers or a group of researchers. It is the most basic of bibliometric techniques. Publication counts are sometimes considered to be a quantification of the peer review process, because each publication is likely to have been approved through a peer review system prior to gaining acceptance for publication. Yet they are more accurately described as a simple measure of scientific productivity or output. Publication counts are often useful for providing a measure of total research outputs. Their primary drawback is that they do not discern the quality of these outputs.

There have been many studies that support publication counts as a reasonable proxy of scientific output (Stephan, and Levin, 1988). Narin found high correlation between bibliometric statistics and non–literature based indices (Narin, 1976). This and other studies have found a strong correlation between publication counts and peer review results (King, 1987; Jones 1980). These studies provide support for the relative accuracy of these bibliometric measurements.

Citation Counts: While publications offer a reasonable proxy for the quantity of research output, citation counts are considered to go one step beyond publication counts and address questions of quality, influence, and the transfer of knowledge. The data being considered are the frequency by which a particular publication, or author, is cited in other publications. Proponents of citation counts as a quality indicator argue that important works will have a number of citation. In this case, citations are viewed as almost a qualitative peer review type of measurement. More specifically, citation counts are considered to assess the impact of a particular scholarly work (Irvine, 1989; Garfield et al., 1978). The assumption is that influential publications are more likely to be cited more frequently.

To illustrate how the bibliometric approach is used in practice to provide an appreciation of how a laboratory might use the method Table 11 shows how 10 large laboratories rank in terms of number of publications, citations, and citations per publications in the area of physical sciences. The reference period for the date was from 1981–1991. The ratio of citations to publications is used a measure of the impact of the publications.

TABLE 11

Top Ten U.S. Research Institutions in the Physical Sciences, 1981–1991, Ranked by Citation Impact

Rank	Name	Papers	Citations	Impact[a]
1	Institute for Advanced Study, Princeton	1,462	25,538	17.47
2	Xerox Corporation	1,619	26,516	16.38
3	AT&T Corporation	10,340	169,031	16.35
4	Harvard University	7,049	110,760	15.71
5	Princeton University	5,593	85,423	15.27
6	University of California, Santa Cruz	1,541	22,963	14.90
7	IBM Corporation	8,929	127,092	14.23
8	University of California, Santa Cruz	4,583	64,744	14.13
9	Caltech (including Jet Propulsion Laboratory)	9,160	128,919	14.07
10	University of Chicago	4,781	65,203	13.64

[a] Total number of citations divided by number of published papers.

SOURCE: Institute for Scientific Information, Science Indicators Database, 1981–1991,Philadelphia.

Co–word Analysis: Co–word analysis is a relatively new addition to bibliometrics. Developed in the early 1980s, co–word analysis is a method by which enables key words to be assigned to papers or articles (Callon, et al., 1983). Papers which include the same key words or sets of words are linked via a mapping technique (Mullins, et al., 1988, Rip and Courtial, 1984). This technique is used as part of scientific mapping. The co–word analysis uses co–occurrences of key words to develop a distance mapping and is usually expressed as a logical tree. The objective is to map national and international cognitive networks in order to position certain research institutes within these networks.

Perhaps the best and also the most controversial use of bibliometric methods to evaluate scientific research is contained in a series of articles by Martin and Irvine (1983a, 1983b, 1984, 1985). They argue claim that citation evidence can be used in conjunction with ex post peer review results to determine productivity of basic research. Their approach involves obtaining counts of publications and citations associated with different research programs in a specific scientific field or topic and then normalizing the outputs from each program with respect to the scale of the research effort, using cost, person hours of effort, or some other readily available input parameters. The normalized research outputs for each program are aggregated to arrive at a single indicator of research productivity enabling research programs to be ranked based on their productivity. Finally, an ex post peer review is conducted for each of the research programs and the citation rankings are compared with the rankings from the peer review. If the productivity analysis is consistent with the peer reviews, Martin and Irvine argue that these "converging partial indicators" can be used as a basis for shifting resources from less efficient research programs to more efficient ones.

In the United States, the earliest studies in bibliometrics were supported by NSF. However, most of the work in this area has been sponsored by the Program Evaluation Branch of NIH. The first group of NIH bibliometric studies, conducted by Grace Carter (1974), analyzed over 800 research grants funded by NIH. She found that grants which were renewed had higher publication rates than did those which were not renewed and that priority scores from peer reviews of grant applications were highly correlated with the number of subsequent publications. More recently, NIH has sponsored bibliometric studies to determine its effectiveness against alternative methods for supporting research and to evaluate manpower training programs (OTA 1986).

Bibliometric methods are considered valuable because they provide a means of measuring the output of a research project or program along several dimensions. Publications, adjusted for the quality of the journal in which they appear, offer a crude measure of the information produced by a research project. Furthermore, Martin and Irvine (1992) have shown that when publication counts can be normalized on the basis of other research inputs, to derive indication of the productivity or efficiency of a research projectd. These measures of research output and efficiency can be compared only in fields where publication practices and incentives are uniform for ranking purposes. Similarly, citation counts can also serve as an indication of the impact of a basic research effort on the information base of a particular field. Taken together, these indicators can be used to help differentiate between research projects, programs or research groups based on their output, impact, and productivity.

However, there are at least three limitations with this approach when used to support research decision–making. First, bibliometric methods can be criticized because measures of output and productivity are too narrow (Cozzen 1992; Brown 1996). This criticism rests on the premise that most R&D activities outputs or objectives tend to be multi–dimensional. Facilities, laboratories and scientific institutions have objectives other than producing publications and receiving citations, such as training and educating future scientists, which are not normally taken into account by bibliometric methods. Measurements that do not recognize these external benefits will understate the productivity of the research being evaluated and under–allocate resources to its purpose. Second, while bibliometric methods may be able to demonstrate a high correlation between peer review results and output, they have no inherent predictive capability. This limits the applicability of these methods to evaluating basic research on an ex ante basis.

Finally, these methods, like peer review approaches, cannot be used to make resource allocation decisions involving tradeoffs between research and non-research activities. Indeed, some critic of this approach (Chubin 1981) contends that structural differences between research fields and disciplines in some cases make it impossible to compare even the most sophisticated bibliometric measures of research output from different research projects.

Evaluation Models for Group II Technologies

Human Capital Method

The methods discussed so far all rely on measures of market activity to evaluate federal R&D. However, in many instances federal R&D is used to improve human welfare in ways that cannot be entirely captured through the use of market prices. This is true of investments in human health where markets do not exist for pricing lives saved or of greater longevity. This is also true, in many cases, of energy R&D investments whose primary benefits are associated with environmental quality where market prices of complementary goods (such as pollution reduction) fail to capture the social value associated with their use, or where there are no market prices reflect environmental values.

One way that economists have attempted to overcome this limitation in the health field is by valuing the impact of R&D on direct and indirect health costs to individuals. The so-called "human capital" approach is based on the assumption that changes in morbidity and mortality can be valued in terms of the opportunity cost of the resources used in treatment and the income foregone that can be indirectly linked to sickness or death. This approach was used by Mushkin (1979) in conjunction with residual imputation to quantify the value of biomedical research during the period 1900–1975. She first calculated the direct costs associated with different chronic illnesses, including expenditures on hospitalization, physicians, drugs, etc. To this were added (1) the morbidity costs due to losses incurred by an individual when illness or disability results in absence, either temporary or permanent, from the work force and (2) mortality costs due to premature death. The latter was estimated by the net present value of an individual's earnings foregone due to premature death. These costs vary according to the occupational, age and sex composition of the population to which they are applied.

Mushkin used a combination of the production function and human capital approach to determine the effect of biomedical research on the reduction of mortality. Indicators of technological change attributable to biomedical research could not be found. Consequently, any reduction in mortality that could not be attributed to other factors was attributed, as a residual, to advances in biomedical research. Using this approach, Mushkin estimated that a one percent increase in biomedical research during the period resulted in a 0.05 percent decrease in mortality. She also estimated biomedical research contributed to about 40 per cent of the reduction in days away from employment because of illness. Finally, Mushkin used the human capital cost estimates to calculate the value of premature deaths avoided and work years gained due to biomedical research. She found that these values, when combined, were approximately $150 billion in present value terms. This was consistent with an annual rate of return on investment of 46 percent.

While the human capital approach is specific to health-related fields, it represents one way to overcome a more general problem associated with the valuation of goods and resources that are not sold in markets (non-market goods). The theoretical advantage of placing a monetary value on the benefits of R&D that improves human health or on other non-market goods and services is that it enables explicit comparisons of the tradeoffs associated with alternative uses of federal funds. However, there are also theoretical problems with the human capital approach. Specifically, the use of expected future earnings as a measure of the value of life can be faulted on four grounds: (1) it implies a positive value for the death of someone whose expected contribution to Gross National Product (GNP) is negative; (2) it ignores the feelings of the potential victims; (3) it assumes that the only contribution that a human life makes to society is to the GNP, and (4) to enable the value of projects to be compared, it requires the implicit assumption that lives saved are of equal value. Finally, as a practical matter, the approach used by Mushkin is primarily oriented toward ex post evaluation and cannot be used on an ex ante basis to make R&D decisions, unless one assumed that the average rate of return on past biomedical R&D equals the marginal rate of return on proposed biomedical research.

The Case Study Method

Case Study Method begins by identifying a successful technological innovation program to study. The history of the development of that particular innovation is then traced backward to its origin. In this process the major research events that are hypothesized to have contributed to the innovation are identified. These research events are also traced back within a designated time horizon and classified according to a spectrum of sponsored research activities. These classifications of contribution are then summed and reported as percentage attributable by each type of activity to the creation of the innovation.

Three criteria are important in the case study methodology: the research question addressed; how cases are selected; and the analytical framework used to evaluate the case. On the surface it often appears that the research question is common in all case studies; (i.e., what is the impact of R&D on the innovation?) Public R&D case studies would probably focus on the impact of R&D on social welfare while private sector studies emphasize the impact on profits or industry competition. In selecting a case to study, adverse selection biases must be avoided or the creditability of the effort will be called into question and not useful for making general statements about the value of programs. Analytical framework, measurement techniques, types of data, and methods of analysis can differ from case to case. Many of these issues are interrelated in retrospective analyses. Thus the collection of data is also considered as a process of describing and interpreting the history of a technology or industry. Second, with the exception of the major government evaluations, the elements that comprise the analytical framework are rarely reported in any detail.

Well-known case studies are: Hindsight (1967), sponsored by the U.S. Department of Defense; Traces (1968), sponsored by the National Science Foundation; and a continuation of the DOD Traces project by Battele (1973).

Hindsight consisted of a cost-benefit assessment of the relative contribution to weapons innovations research directed towards a specific technological goal versus non-directed or basic research. There were two steps in the analysis. The first step was to estimate costs of developing current weapons innovation and compare those to the costs of operating the weapons system that it replaced. The second step was to classify research events in two categories (Sherwin & Isenson, 1967):

1) Science events – theoretical as experimental studies of new or unexplored phenomenon. Science events were further defined as either undirected science, where the object of the work is the advancement of knowledge, or directed science, where the objective of the work is to produce specific knowledge for a particular use.

2) Technology events... the conceptualization or demonstration of the possibility of performing specific elementary function with the use of new or untested concepts, principles, techniques, or materials... ... and the development of new manufacturing techniques," (Sherwin & Isenson, 1967). The expert of the cost-benefit framework was to treat basic research as competing with applied and development research, as opposed to complementing those activities, by assessing the relative contribution of science and technological events (Mowery & Rosenberg, 1982).

The development of a weapons-related technology was traced back 20 years in Project Hindsight. The research events were also traced back within that 20-year time frame. Restricting output to this time frame was justified on the grounds that each generation of weapons technology lasted roughly 20 years and the study was designed to assess the contribution of current research (Sherwin & Isenson, 1967).

Traces and the Battelle study employed a different classification scheme and extended the time horizon to 50 years. The research events were grouped according to three categories (IITRI, 1968):

1) Non-Mission Research - Motivated by interests in increasing knowledge and scientific understanding without regards to application.

2) Mission-Oriented Research - Performed to develop information for a specific application concept prior to development of a prototype product or engineering design.

3) Development and Application - Inventory prototype development and engineering directed toward the demonstration of a specific product or process of marketing.

A rich variety of methods are brought together by government agencies conducting case studies of research program. Three methodological combinations seem to dominate. First is a combination of retrospective analysis and aggregate statistics. A history is developed of the agency's support for a particular technology innovation program by describing the projects the program has sponsored. This history usually include a description of how the agency became interested in the particular program, a review of the amount of funds that have gone into supporting the program, and the involvement by other organization in the program. This history is compiled by the sponsoring agency either directly or on a contract basis through telephone interviews with program participants. The findings from these interviews are presented descriptively. An example of this type of study is found in evaluation of technology transfer– conducted by Oak Ridge National Laboratory (Brown, et al., 1987; Brown, et al., 1989).

A second form of impact analysis included in case studies brings together retrospective analysis with a production function. This approach is used primarily when evaluating whether projects had an impact on industry operations, such as the National Institute of Standards and Technology (NIST) goal of increasing productivity (Charles River Associates, 1981). Here, the retrospective analysis chronicles the contribution by the agency to the development of a particular technology. However, the goal is to develop impact measures for the case under study (Logsdon & Rubin, 1985). For example, the research instrument used to develop these measures at NIST was a mail questionnaire sent to project managers in participating industries (Charles River Associates, 1981). The production function is used to measure the influence of a technology on the performance of client industries by introducing a measures of technological knowledge as an additional input (Logsdun & Rubin, 1985).

Finally, case study include a form of impact analysis which combines retrospective analyses with peer review. This method has been developed by the Office of Naval Research (ONR) emphasizing peer review as the primary evaluation method (Kostoff, 1988). But the retrospective analysis is extensive, combining tracings of the technology to a specific project

(Kostoff, 1988). Program managers at ONR are responsible for developing materials used in the case study as a part of presentations made before peer review panels. These panels are then given a scoring instrument in which element of the retrospective analysis are evaluated along with data on other technical and scientific dimensions (Kastoff, 88).

Economic Surplus Method

An important drawback with most of the methods discussed above is that they ignore the fact that technological discoveries can cause changes in the market prices of related goods. These price changes render it difficult to accurately estimate R&D benefits. For example, the so-called "Green Revolution" made it possible to grow high-yielding grain varieties in a number of different parts of the world.

This had an important consequence in lower world grain prices. The primary beneficiaries of the Green Revolution have been consumers worldwide, who now have access to much lower priced grain, and some producers in developing and lesser developed countries where lower production costs compensated for the decline in prices. Grain producers in the U.S., on the other hand, were adversely affected because they derived none of the benefits of the lower production costs afforded by the Green Revolution, but had to sell their grain on a world market in which prices were depressed, due in part to the higher production made possible elsewhere. To properly value the benefits of research related to the Green Revolution, a method for measuring benefits that takes into account the conflicting impact of research-induced price changes is needed.

The economic surplus approach attempts to accomplish this in two ways. First, the benefit measures that are used by this approach take into recognition the fact that consumers derive benefits from the consumption of a good when they are able to purchase the good at a price less than the maximum amount they would be willing to pay. By the same token, a firm benefits by the production of a good when it is able to sell the good at a price greater than the minimum amount it costs to produce the good. In both of these cases, there is an economic surplus present. The concept of economic surplus and its utility for measuring the benefits of R&D can be dealt with more fully; for this report, what is important about this concept is that it provides a theoretical basis for explaining benefits to consumers and producers in common units. Furthermore, it does this in a way that allows one to take into account the sometimes uneven impact of R&D–induced price changes on the benefits of both groups.

A second aspect of the economic surplus approach is that it uses mathematical representations (i.e., models) of supply and demand curves in relevant markets as the basis for measuring these surplus changes. With appropriate data, parameters of these supply and demand curves can be empirically estimated and the models linked to simulate the economic behavior of buyers and sellers in relevant markets. The economic surpluses of consumers and producers can be calculated from information obtained directly from the supply and demand curves. R&D investments are generally modeled using a production function approach, such that a simulated increase in R&D funding lowers the marginal cost of producing relevant goods over a substantial range of output and makes these goods more attractive to consumers. The models then simulate the process of exchange with the new technology in place until market equilibrium is restored and the consumer and producer surplus calculations are repeated. The periodic benefits of the R&D investment are calculated as the change in total economic surplus.

The best examples of this approach are to be found in the ex post evaluation of R&D in the agricultural sector. The first such major use of this approach was by Schultz (1953) who calculated the value of the inputs saved in agriculture due to adoption of improved, more efficient production techniques. Following Schultz, Griliches (1958) used this approach to estimate the loss in surplus to consumers that would occur if research on hybrid corn had not occurred. These early studies were methodologically flawed because of overly simplistic assumptions made by Schultz, who assumed that individual demands for agricultural commodities were not price-sensitive, and by Griliches, who alternately assumed that the supply of corn was totally price-insensitive or else that the price of corn was insensitive to changes in supply. Peterson (1967) relaxed these assumptions in his study on poultry research in the U.S. and calculated the effect of this type of R&D on consumer and producer surpluses. He then compared these benefits with the costs of the R&D and estimated a rate of return of about 25 percent on this investment. Peterson's work is generally regarded as the standard against which methodological improvements are measured.

The economic surplus approach is general enough that it can be applied broadly to R&D impacts on both market and non-market goods, although some of the methods for estimating demand for non-market goods are highly controversial. In either setting, however, this approach does have several important limitations. First, the data requirements necessary to estimate demand and supply curves are generally not available from published sources and are often difficult to obtain either for cost or proprietary reasons. Second, while spillovers into other markets can be modeled using this approach, the data limitations, which are already severe, become much more serious. Third, this approach generally does not work well with basic research since it is extremely difficult to trace the effects of basic research to all of the goods that have been influenced by it. Finally, this approach can only be used fruitfully in an ex post evaluation setting. Uncertainties about the effects of basic research on the supply and demand functions of both market and non-market goods, coupled with its data intensiveness, make it a poor candidate for ex ante evaluation.

Production Function Method

This technique represents the traditional approach to measuring the impact of R&D on productivity. It shares with other methods the idea that R&D can be described in terms of a production process. The process is characterized by a technology that transforms inputs, such as capital, and labor, into outputs. In this particular framework, R&D is treated as an input to production, while the outputs of the R&D process can include new information, changes in product quality, or new technologies.

Although the description of the production process and, particularly its outputs, are somewhat abstract, the application of the production function method is more straightforward. In practice, an R&D production function is postulate that relates the output of goods and services by a particular sector to observable inputs, including R&D. Multiple regression analysis is then used to derive empirical estimates of the parameters of the production function. The form of the production function is chosen such that the regression model is then used to evaluate the effect of changes in R&D on the output of goods and services in the industry. This measure, marginal R&D productivity, is then used to derive estimates of the rate of return to the R&D.

This method has been employed in an aggregate economic and in less aggregate industry settings. The results have been mixed. Griliches and Lichtenberg (1984) estimated a 1.5 percent average rate of return to federal R&D in 27 industries for the period 1959–1976. The corresponding rate of return to private R&D for these industries was almost 22 percent. Much higher rates of return to federal R&D have been found, particularly in agriculture where rates of return have typically been estimated in excess of 25 percent (Evenson et al. 1979), well above any reasonable measure of the opportunity cost of capital.

The low rates of return to R&D, found by Griliches and Lichtenberg, have been rationalized by Terleckyj (1974) as consistent with the behavior of a firm that is given a free good (R&D in this case) to use in production: the firm will use the good up to the point where additional amounts do not produce additional outputs or to the point where its marginal product is zero. A second explanation for apparently low rates of return to R&D has been advanced by Mansfield (1984) who contends that federal R&D in many cases does not contribute directly to output growth, but instead tends to enhance the profitability of private R&D. A third explanation is that a firm's rate of return to R&D is small because of R&D–induced decreases in the prices of goods and services. The benefits to consumers of this R&D, on the other hand, may be substantial, due to the same low prices. However, the production function approach limits the benefits from R&D to firms and generally does not measure benefits to consumers.

Particularly in agriculture, the production function approach has proven useful at isolating the effect of R&D on productivity. However, this approach suffers from a number of theoretical and practical limitations. From a theoretical prospective it is not clear, as Mansfield (1984) suggests, that R&D investment should enter the production function in the same form as other inputs such as labor and capital. A second theoretical limitation is that this method does not account for spillover consequences of R&D on other industries, nor the benefits that R&D may generate by lowering the cost of goods and services to other firms and consumers. Finally, there is the question of how the results of productivity analysis can be disaggregated on an ex ante basis to compare individual R&D technologies. The application of productivity analysis to evaluate output of basic research is particularly problematic due to unknown or unintended effects. On a broader scale, productivity analysis may provide some indication of which industries should receive continued support based on higher-than-average private market returns; however, it is not at all useful in deciding whether federal agencies should allocate resources to individual programs.

Residual Imputation Method

A traditional method for valuing the productivity of inputs that are hard to price or measure is the residual imputation approach. In this approach, budgets delineating resource usage are constructed for representative programs being investigated. These budgets are used to determine the costs[2] and quantities associated with each input used in production. These costs are summed to determine what it will cost to produce a unit of output at an appropriate scale of operation. This normally includes a benefit margin, which are figured as a payments by the program to participants. Next, the marginal output of the program is valued at its market price and the costs associated with this additional unit of output are subtracted. The residual is then used to value the contribution of the unobserved, or hard to price, input to the value of output.

This approach has traditionally been employed as a method of determining the social value of inputs, such as water from irrigation projects, which are not priced in competitive markets at their true opportunity cost. An interesting and less rigorous (but conceptually similar) approach has been used by Mowery (1985) to quantify the benefits associated with federal R&D

---

[2] The appropriate costs used in the budgets should be opportunity costs (what is being foregone by using that input) and not its historical cost.

support of commercial aviation technology. In that study, a relatively simple index of aircraft performance (number of available seats multiplied by air speed) to reflect changes in aircraft technology between 1940 and 1983. He then calculated the change in direct operating cost per passenger mile for the same period.

Mowery combined these two indices to compute the cost to society of using the 1940 technology to carry the volume of passenger traffic in the U.S. in 1983 would have been roughly $25 billion, as opposed to the actual cost of transporting this traffic which was about $6 billion. Thus, by imputing all of the increase in productivity and all of the decrease in cost to federal R&D, Mowery suggests that the $19 billion annual saving to society can be used to approximate the total benefits produced by federal R&D in this area. According to Mowery, this translates into a rate of return on federal R&D of about 24 percent, again substantially higher than the opportunity cost of capital in most alternative private investments.

This approach is best suited as a preliminary method for evaluating R&D. It has virtually no ex ante application; it is not well–suited for valuing the benefits of a technology to consumers due to lower product prices; and it does not include spillover effects on other industries nor environmental effects. Moreover, the residual value that is imputed to R&D could just as easily come from other sources that are equally hard to observe and/or measure. However, in the absence of a great deal of data, approaches like those used by Mowery can prove useful in screening federal R&D programs for further, more in–depth reviews.

Accounting Methods

Accounting methods are among the simplest methods to apply and most frequently used by firms in the private sector and, occasionally, by federal agencies, to measure the value of

R&D. The measures used most often include net present value (sometimes referred to as discounted cash flow), internal rate of return, and project payback period. This method varies from one application to another, but generally involves the following process. First, the expected (for ex ante analysis) or observed effect (for ex post analysis) of a technology on net revenues of a firm or market sector is calculated over time. This estimated flow of future benefits is then discounted over time to derive estimates of their present values to reflect the alternative earning opportunities of the R&D investment. Summing these discounted values over time yields the present value of net benefits attending the project. The same procedure is applied to R&D costs to determine the present value of project costs. The net present value of the R&D investment is determined by subtracting the present value of project costs from the present value of the project's benefits to arrive at net benefits. The internal rate of return is the discount rate that is necessary to equates the present value of the project's net benefits to project costs. Finally, the payback period of the investment is calculated as the amount of time required for the present value of the benefits to equal the present value of project costs. Repeating this calculation for all technologies within a program and summing enables one to derive a rate of return for the overall program portfolio.

As stated above, accounting methods are used more frequently in private industry for selecting R&D projects than in the federal sector. One example of the use of accounting methods to perform ex ante project selection involved the screening of energy conservation programs by the Energy Development and Research Administration (ERDA) and later DOE (Roessner 1981). Project selection models were used to calculate payoffs associated with research investments on different technological strategies for conserving energy.

While the nature of inputs and outputs of these models are varied, energy cost savings per barrel of oil, internal rate of return and length of project payback period can be used to compare investments. By providing a common set of metrics, the accounting method can be used as an ex ante basis to compare investments in R&D with other forms of investment. Accounting methods are most useful in measuring the net benefits of an R&D project objectives when: (1) the project missions are narrowly focused on making an incremental improvement over conventional technology or reducing its cost; (2) the project does not substantially influence the market price of the technology; (3) the market for the technology is competitive, and (4) there are no spillovers, benefits or costs, into other markets. Under these circumstances, accounting methods can be used to approximate the benefits to society of the R&D investment. The fact that these conditions are rarely met in R&D investments in basic research makes this type of approach particularly inappropriate for evaluating that kind of federal investment.

User Evaluations

From the standpoint of program evaluation, the key question in judging importance is who does the assessing. Next–stage users are often involved in this judgment. When importance is judged with regard to bodies of scientific knowledge, researchers must judge that quality – but not the researchers supported by the program, nor those who choose the projects it support. Instead, next–stage users are researchers outside the program, in the areas where the program's work is claimed to have an impact. Agencies that create generic knowledge resources and human capital, as discussed earlier, can in addition identify stakeholder groups for the resources they produce – that is, groups that use the bodies of knowledge and talent pools that the agencies develop, although not the immediate knowledge outputs of specific projects. Such groups can be involved in detailed program evaluation processes.

In mission-oriented programs, next-stage users work in the areas of practice where knowledge developed out of fundamental research is intended to be serve. Thus, it is quite common to find industrial representatives on applied research program evaluation teams; ONR involves DOD technology transfer agents; and the Agricultural Research Service invites large farmers to its evaluation workshops. ARL includes soldiers who would work with the weapons being developed - in its strategic planning process, opening the door to the inclusion of other end users in research management processes elsewhere. The Office of Industrial Technology (OIT) has extensive industry representation on its advisory boards.

The state of the art in research program evaluation has not developed effective ways to translate descriptive knowledge that users bring to the program evaluation process into performance indicators. Nor has it needed to, since users could be involved alongside technical reviewers in any fully developed program evaluation. Generally, however, next-stage users may need to be treated as the "customers" of a research program and surveyed for their satisfaction. This would be a step toward evaluating the results, rather than merely the activity, of research. Appropriate survey instruments and samples could undoubtedly be developed. The Army Research Laboratory, for example, includes customer satisfaction ratings in its summary performance indicators, gathering them on a simple customer feedback form sent out with all final project results.

It is well to keep in mind, however, that there are conflict of interest problems in user ratings of research programs. Next-stage users are the recipients of a free service provided by the federal government, and have a stake in expressing high satisfaction with the programs that benefit them, without regard to their efficiency.

Research programs consisting of technologies for which output is public are also generally expected to contribute to certain broad, federal goals, even when these are not listed as among the program's specific objectives. Indicators of performance in relation to these goals should also be included for program evaluation purposes, unless they are not applicable in a particular case (for example, undergraduate training goals in relation to a national laboratory's research program). In theory, indicators on these criteria could be included in performance plans and reports as well. Examples of such indicators appear in Table 12. All the items on the list represent output indicators, which could be gathered from principal investigators at the completion of research projects and aggregated at agency level. Partnership indicators can also be gathered from the published literature. If such data were collected in final project reports, however, it would be important to communicate the portfolio concept clearly to both investigators and program managers. When the projects are gathered into a portfolio, not every project needs to produce each of the outputs on the list, even though in the aggregate, fundamental science agencies expect to create desired outcomes through these routes. To convey any other message would be to limit the flexibility needed for creative work.

Program managers and participants often perceive the most important characteristics of the knowledge produced by research programs in terms of factors that go beyond both quantity and quality. In disciplinary programs, the theoretical significance of the knowledge is frequently the paramount consideration. Have the researchers in the program enriched the whole field through their insights? Have they developed concepts, methods, or models that apply widely?

$$Q = Q(x,y,z) \quad [1]$$

Where Q is output or outcomes of a basic research program, such as the Office of Basic Energy Sciences, designed to achieve the following mission:

x = excellence in education

y = new knowledge about energy that is useful to society

z = respected leadership in science

Achieving excellence in education, providing new knowledge of service to society, or attaining a leadership position in science requires the use of resources that are made scarce by amount of budget allocations from Congress. The production resources can be subdivided into the traditional categories of factor inputs: labor and capital. Therefore, we might define the resources of the Office of Basic Energy Sciences as follows:

$$x = x(L_x, K_x) \quad [2]$$

$$y = y(L_y, K_y) \quad [3]$$

$$z = z(L_z, K_z) \quad [4]$$

constrained by

$$L_x + L_y + L_z = L \quad [5]$$

$$K_x + K_y + K_z = K \quad [6]$$

where (L_i = x,y,z) and K_i(i = x,y,z) represent amounts of the agency's human and capital, respectively, that is available for achieving the three sub-goals.

If the objective included in the strategic plan of the agency is to maximize objective function [1] with respect to each of the separate arguments in its strategic objective [2–4], this can be summarized as following:

$$dQ = \frac{\partial Q}{\partial x}dx + \frac{\partial Q}{\partial y}dy + \frac{\partial Q}{\partial z}dz = 0 \qquad \text{[6a]}$$

But since the achievement of x, y, and z depends on the amount of human and capital resources dedicated to accomplish these three separate components of the agency's overall strategic plans and each sub–goal contributes directly to the agency's overall mission; then each category of the agency's strategic goal must be maximized with respect to each of its arguments. That is:

$$dx = \frac{\partial x}{\partial L_x}dL_x + \frac{\partial x}{\partial K_y}dK_y = 0 \qquad \text{[6b]}$$

$$dy = \frac{\partial y}{\partial L_y}dL_y + \frac{\partial y}{\partial K_y}dK_y = 0 \qquad \text{[6c]}$$

$$dz = \frac{\partial z}{\partial L_z}dL_z + \frac{\partial z}{\partial k_z}dK_z = 0 \qquad \text{[6d]}$$

substituting [6b] –[6d] into [6a], the conditions that the objective function is maximized without constraints is as follows:

$$Q'(x)\left[\frac{\partial x}{\partial L_x}dL_x + \frac{\partial x}{\partial K_x}dK_x\right] + Q'(y)\left[\frac{\partial y}{\partial L_y}dL_y + \frac{\partial y}{\partial k_y}dK_y\right] + Q'(z)\left[\frac{\partial z}{\partial L_z}dL_z + \frac{\partial z}{\partial K_z}dK_z\right] = 0$$

The general solution to the efficient allocation of labor and capital across different strategic goals can be accomplished by manipulating the above equation:

$$\begin{aligned} Q'(y)\; y'(L_y) &= Q'(x)x'(L_x) = Q'(z)\; z'(L_z) \qquad \text{and} \\ Q'(y)\; y'(K_y) &= Q'(x)x'(K_x) = Q'(z)\; z'(K_z) \end{aligned}$$

The above condition assures that resources are allocated efficiently such that their respective contributions to the strategic objective, y, when weighted by the contribution of y to the agency's overall objective must be equal to the contribution of labor and capital to the accomplishment of strategic goal x weighted by the contribution of x to the overall objective of the agency.

Evaluation methods are generally retrospective and, therefore, rely upon ex–post evaluations of R&D Program Performance. Resource allocation, however, requires federal agencies, i.e., DOE to predict future accomplishments by program area and budget resources accordingly. Perhaps a more efficient method is to gather data through peer review and use expost evaluation data to predict accomplishments for a particular research portfolio. For example, categories to predict ex–ante outcomes are based upon an independent assessment of projects at the time they are added to a program's portfolio. It seems plausable to rely data collected on completed projects to identify attributes in proposals that are useful at predicting success. To avoid sample bias, the same type of measures must be collected on all projects in a portfolio. One could use analytical results from previous assessments for funded projects to approximate future performance for technologies contained in the current inventory. The model that relies on ex–post data could then be used to estimate value of parameters necessary to quantify success for existing technologies. The procedure could be replicated in other programs to provide comparison. The model might be implicitly described as follows:

$$y_i = f\,(x_i \ldots x_n)$$

Where xi represent numerical scores assigned by external review teams to a specific criteria on that are part of a portfolio of research activities included in a program. Y_i are ex-ante assessments of program effectiveness after completion.

Regression analysis should be used to determine how efficiently a set of projects selection variables predict the ex-ante assessment on the y output measures. Under the assumption that the parameters are stable over time and across technologies, parameter derived by estimating the influence of the ex-post values of the x's on the ex-ante value of the y's can then be used to develop expectation for each program area. These parameter could serve as productivity indexes that isolate the influence scores assigned by external peer review of each assessment category on future accomplishments for the project, program.

To implement the model each project that receive DOE funding must be evaluated according to the same criteria. Weights could then be assigned to each of the projects reflecting their predictive value at achieving the broader goals of the agency.

No model will provide 100 percent accurate. That is, observed outcomes of project investments in the future will not be possible to predict. Thus when DOE presents its expected outcome, ranges must be established on those estimates that reflect the level of confidence that are provided by parameter developed from the model. This would present the agency with statistical measures that better lend themselves to quantification.

This model would rely upon external peer reviews to accomplish several objectives. First, quality control area the project selection process would leave program assessment to the scientific community. This lends some degree of credibility to the portfolio of research project included in the inventory. The assessment of the usefulness of the project and they are completed is again left to the experts. Then by modeling the linkage between what experts said on the front end this ex–post assessments, parameters could be useful at developing a list of characteristics look for in the portfolio or projects that are favorable at enhancing the likelihood that the projects will prove successful on the end.

CONCLUSIONS

From this discussion, it should be clear that the methods available for assessing performance of basic research programs may be quite reasonable to use in the context of program evaluation, where multiple indicators are the rule and knowledgeable people are available to integrate them wisely into an assessment. Cautions and caveats about such use have been discussed in the preceding subsections, and are already embodied in the practice of research program evaluation, particularly in the use of multiple indicators and their combination with technical review. In efforts to develop a few summary indicators, however, agencies will apparently need to pare down this full data set to its essential elements. A different set of cautions applies in this situation.

A frequently voiced fear is that program evaluations will encourage agencies to measure what is easy and ignore that which is important. One can easily picture the indicators that would fill this description and satisfy possible administrative requirements for a limited, objective, quantitative set, with which one could set baselines and compare later performance:

- publication counts (year of review) (x y, z)
- citations per publication (lagged three years; compared with average for journals 1 x where they were published)
- doctorates produced (x)

- entering research careers
- entering careers in practice
- undergraduate students involved
- user involvement and satisfaction ratings (in–science users for some programs, outside–science users for others)

The problem with the set, of course, is that it omits virtually all of what researchers themselves consider important about their work. One could have an agency full of programs that performed beautifully according to these indicators, and still be at the trailing edge of every scientific frontier.

The key to conducting effective program evaluation, therefore, lies not in the indicators themselves, but in the larger effort in program evaluation in which they are embedded. The indicators, preferably reported in context as the Swedish example above illustrates, can provide a bare–bones description of whether the program is producing the basic expected outputs, and can point toward programs that are particularly in need of evaluation. But the more detailed information that is needed for general program planning and resource allocation, including descriptive judgments and analysis, still needs to come from the more intensive and interactive process of detailed program evaluation.

Many of the key issues with regard to program evaluation, however, lie outside the control of agencies, and in the hands of those who receive and use the performance measures. Optimists claim that GPRA will revolutionize government management by focusing program attention intelligently and diligently on results. Pessimists fear that it will create busy work number–generating, then put a simple–minded tool in the hands of decision makers who already pay too little attention to the programs they expand, cut, and re–arrange.

Where the actual result falls--probably somewhere between the extremes the optimists and pessimists describe--will depend first on what the Office of Management and Budget encourages and requires of agencies as it collates their responses into government-wide performance plans and reports, and second on how the indicators are used in Congress. The first set of results will not be in Congressional hands until March, 2000. If the election trends of the early 1990's continue, most members of that future Congress have not yet been elected, and therefore probably have not yet begun thinking about how they will react to the indicators the research community is now beginning to prepare for their perusal.

REFERENCES

Arrow, K. 1962. "Economic Welfare and the Allocation of Resources to Invention." In The Rate and Direction of Inventive Activity: Economic and Social Factors, pp. 609–626.

National Bureau of Economic Research, Princeton, New Jersey.

Bator, F.M. 1958. "Anatomy of Market Failure." Quarterly Journal of Economics 72(3): 351–379.

Bush, V. 1945. Science. the Endless Frontier. U.S. Government Printing Office, Washington, D.C.

Carter, G.M. 1974. Peer Review. Citations. and Biomedical Research Policy: NIH Grants to Medical School Faculty. R–1583–HEW,Department of Health Education and Welfare. Washington, D.C.

Chubin, D.E. 1976. "The Conceptualization of Scientific Specialties." The Sociological Quarterly 17(4):448–476.

Chubin, D.E. 1981. "Constructing and Reconstructing Scientific Reality: A Meta–Analysis." International Society for the Sociology of Science Newsletter 7(3):22–28.

Evenson, R.E., P.E. Waggoner, and V.W. Ruttan. 1979. "Economic Benefits From Research: An Example From Agriculture" Science 205(4411):1101–1107.

Finneran, K. 1986. The Federal Role in Research and Development, Report of a Workshop. National Academy of Sciences. National Academy Press, Washington, D.C.

Griliches, Z. 1958. "Research Costs and Social Returns: Hybrid Corn and Related Innovations." Journal of Political Economy 66(4):419–431.

Griliches, Z. and F. Lichtenberg. 1984. "R&D and Productivity Growth at the Firm Level: Is There Still a Relationship". In R&D, Patents. Productivity and Growth, ed. Z. Griliches. University of Chicago Press, Chicago, Illinois

Hueth D.L., R.E. Just, and A. Schmitz. 1982. Applied Welfare Economics. Prentice Hall, Inc., Englewood Cliffs, New Jersey.

Kuhn, T.S. 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, Illinois.

Logsdon, J.M. and C.B. Rubin. 1985. An Overview of Federal Research and Evaluation Activities. Program of Policy Studies in Science and Technology. George Washington University, Washington, D.C.

Mansfield, E. 1984. "R&D and Innovation: Some Empirical Findings." In R&D Patents. Productivity. and Growth, ed. Z. Griliches. University of Chicago Press. Chicago, Illinois.

Martin, B.R. and J. Irvine. 1983a. "Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy." Research Policy 14(1):61-90.

Martin, B.R. and J. Irvine. 1983b. "Assessing Basic Research." Social Studies of Science 13(3):49-86.

Martin, B.R. and J. Irvine. 1984. "CERN: Past Performance and Future Prospects." Research Police 13(1-3):183-210;247-284;311-342.

Martin, B.R. and J. Irvine. 1985. "Quantitative Science Policy Research." Paper presented to the Task Force on Science Policy of the Committee on Science and Technology, U.S. House of Representatives.

Mowery, D.C. 1985. "Federal Funding of R&D in Transportation: The Case of Aviation." Paper presented to the National Academy of Sciences workshop on The Federal Role in Research and Development, November 21-22, 1985. National Academy Press, Washington, D.C.

Mushkin, S. 1979. Biomedical Research: Costs and Benefits. Ballinger Publishing Co., Cambridge, Massachusetts.

Office of Technology Assessment (OTA). 1986. Research Funding as an Investment: Can We Measure the Returns. Congress of the Unites States, Office of Technology Assessment, Washington, D.C.

Peterson, W.L. 1967. "Returns to Poultry Research in the United States." Journal of Farm Economics 49(3):656-669.

Roessner, J.D. 1981. R&D Selection Models in the U.S. Department of Energy. Georgia Institute of Technology, Atlanta, Georgia.

Rosenburg, N. 1985. "An Historical Overview of the Evolution of Federal Investment in R&D Since World War II". Paper presented to the National Academy of Sciences workshop on The Federal Role in Research and Development, November 21-22, 1985. National Academy Press, Washington, D.C.

Schultz, T.W. 1953. The Economic Organization of Agriculture. McGraw Hill Publishing Co., New York, New York.

Terleckyj, N.E. 1974. Effects of R&D on the Productivity Growth of Industries: An Exploratory Study. National Planning Association Report No.140, National Planning Association, Washington, D.C.

U.S. Department of Energy (DOE). 1982. An Assessment of the Basic Energy Sciences Program. DOE/ER-0123,Office of Energy Research, U.S. Department of Energy, Washington, D.C.

A-1

DESCRIPTION OF SELECTED APPLIED TECHNOLOGIES

Technology	Description	Benefits
Fluorescent lamp electronic ballasts	DOE developed the electronic fluorescent lighting ballast at its Lawrence Berkeley Laboratory in the mid 1970s. The ballast eliminates the flicker and hum of traditional magnetic ballast and saves energy.	Ballast has improved lighting quality and saved consumers $750 million in energy bills.
Software for building design	DOE developed a software tool, DOE-2, that estimates, on the basis of a building's characteristics, its energy use and cost.	Use of this software accounts for $1.9 billion and energy savings for buildings constructed through 1993.
Nickel metal hydride batteries	DOE is supporting the development of a low cost high performance battery for electric vehicles through the United States Advanced Battery Consortium. This battery will double an electric vehicle's driving range between recharges and significantly increased power.	The mandates for electric vehicles in California and in the Northeastern states will create a $350 million market for this battery in 2003.

Technology	Description	Benefits
AC electric drive train	Under a cost–shared contract with DOE, Ford Motor Company and General Electric developed a new electric drive train for electric vehicles that run on AC current.	New design will reduce consumers' cost and allow electric vehicles to enter the market sooner. California laws mandating zero–emission vehicles will result in approximately $70 million in electric vehicle sales in 1998, growing to $350 million by the year 2003.
Electrochemical dezincing of steel scrap	DOE has developed an electrochemical method that removes the zinc from steel scrap so that the scrap can be used in steel making operations.	Electrochemical method will (1) increase production yields and quality and (2) by the year 2000, will save 50 trillion BTUs and reduce raw material imports by at least 75,000 tons of zinc per year, thereby saving $77 million annually.
Integrated gasification combined cycle (IGCC)	IGCC is an advanced coal–burning system that DOE believes will be the power plant of the 21st century.	IGCC technology will (1) reduce sulfur dioxide and nitrous oxide emissions to less than 10 percent of new source performance standards, (2) reduce carbon dioxide emissions by 35 percent to 45 percent, (3) reduce solid wastes by 40 percent to 50 percent, and (4) be less costly to build.

Technology	Description	Benefits
Photovoltaics	Photovoltaics are devices that convert light into electricity. DOEs photovoltaic program has succeeded in reducing the cost of such electricity from 90 cents per kilowatt hour.	$100 million in photovoltaic sales supports or creates 3,800 U.S. jobs.
Mudpulse telemetry	In the 1970s, DOE helped a private company develop an instrument for measuring while drilling that significantly cut the cost and time of drilling oil and gas wells.	Mudpulse telemetry has gained wide acceptance in the drilling industry, and DOE estimates that it has saved the natural gas and oil industry at least $1 billion over the past 20 years.
Carbon dioxide sand fracture production technology	DOE developed, tested, and helped commercialize this technology for stimulating production from natural gas wells. It has been shown to increase production by 200 to 500 percent.	This technology could generate $20 million more revenue over the productive life of some wells.
Atmospheric fluidized bed coal combustor	DOE helped develop a coal combustor that uses low–polluting coal to produce electricity. According to DOE, every U.S. bottler manufacturer now sells a fluidized bed coal combustor.	Over the last 8 to 10 years, more than $6 billion in domestic sales and $2 billion in foreign sales have been reported. Domestic sales alone translate into more than 250,000 jobs.

Technology	Description	Benefits
High–efficiency refrigerator compressors	From 1976 through 1980, DOE sponsored a project that resulted in a 44 percent improvement over the compressor technology used in refrigerators at that time.	Use of the improved compressors pioneered by this research saved consumers at least $6 billion in energy costs from 1980 through 1990.
Flame retention head oil burner	In the early 1970s, a DOE field test established the energy conservation benefits of a new flame retention head oil burner. DOE later published its findings in a consumer–oriented information booklet. Within several years, the flame retention head burner dominated the market for new and replacement oil burners.	Consumers' energy cost savings to date from this innovation total more than $5 billion.