*Jointly published by Elsevier Science B. V., Amsterdam and Akadémiai Kiadó, Budapest*

*Scientometrics, Vol. 34, No. 2 (1995) 163-206*

**Review**

# FEDERAL RESEARCH IMPACT ASSESSMENT: AXIOMS, APPROACHES, APPLICATIONS*

R. N. KOSTOFF

*Office of Naval Research, 800 N. Quincy St., Arlington, VA 22217 (USA)*

(Received April 4, 1995)

## Outline of paper



* The views in this paper are solely those of the author and do not represent the views of the Department of the Navy.

*0138-9130/95/US $ 9.50*

This paper describes the practice of Federal research impact assessment. Evaluation of research impact is described for three cases: Research selection, where the work has not yet been performed; Research review, where work and results are ongoing; and Ex-post research assessment, where research has been completed and results can be tracked. Retrospective methods (such as projects Hindsight an TRACES), qualitative methods (such as peer review), and quantitative methods (such as cost-benefit analysis and bibliometrics) are described. While peer review in its broadest sense is the most widely used method in research selection, review, and ex-post assessment, it has its deficiencies, and there is no single method which provides a complete impact evaluation.

## I. Background

### *A. Definitions*

Research is the pursuit and production of knowledge by the scientific method. Research Productivity is the generation of tangible and intangible products from research. Research Efficiency is the productivity of research per unit of input resource. Research Impact is the change effected on society due to the research product. Research Effectiveness is a measure of the focus of impact on desired goals.

### *B. Increased interest in research impact measurement*

In research sponsoring organizations, the selection and continuation of research programs must be made on the basis of outstanding science and potential contribution to the organization's mission. Recently, there have been increasing pressures to link science and technology programs and goals even more closely and clearly to organizational as well as broader societal goals. This is reflected in a number of studies,[1-3] in the controversial National Institutes of Health strategic planning process, in the controversial statements by the previous National Science Foundation director about closer alignment with industry and other government agencies, and in conversations with numerous government officials.

In tandem with the pressures for more strategic research goals are motivations to increase research assessments and reporting requirements to insure that the increasingly strategic research goals are being pursued by proposed and existing research programs. The 1992 Congressional Task Force report on the health of research[1] stated, as one of its two recommendations: "Integrate performance assessment mechanisms into the research process using legislative mandates and other measures, to help measure the effectiveness of Federally funded research programs".

According to the statement of Genevieve Knezo, a Congressional Research Service representative, at a 1993 research assessment colloquium, "The House Science Committee has asked the Congressional Research Service to develop some options for legislative language that might be included in the mandates of the agencies that they have responsibility for, which would require or in some way discuss the need for R&D to be evaluated. We are exploring that right now. We have a task force in Congressional Research Service composed of about twelve people who are surveying the agencies for which the Committee has responsibility. We are also surveying agencies outside the jurisdiction of the House Science Committee, DOD and NIH specifically".

The Government Results and Performance Act of 1993 (Public Law 103-62) was passed on August 3, 1993. This Act provides for the establishment of strategic planning and performance measurement in the Federal government, and for other purposes. Not only will the Federal agencies be required to establish performance goals for program activities, but as the law states, they will be required to establish performance indicators to be used in measuring or assessing the relevant outputs, service levels, and outcomes of each program activity. Early meetings of the newly-established Federal Research Assessment Network focused in part on the implications for research from the Act.

In November, 1994, the OSTP held a workshop on goals, returns, and evaluation of Federal research. A practitioners' working group on research evaluation was assembled by Professor Susan E. Cozzens to write a concept paper on research program evaluation that was available to OSTP in planning the fall workshop. The concept paper addressed three issues: (1) The definition of research program evaluation; (2) What are OSTP's options for guidance to agencies on evaluating their research programs; (3) What are the cautions, caveats, and needs for further research in this area. The author participated in the working group.

Due to increased world competition, and the trends toward corporate downsizing, parallel pressures exist for industrial research organizations to link research programs more closely with strategic corporate goals and to increase research performance and productivity. In tandem with the increasing governmental interests in research assessment stated above, there is considerable industrial interest in research assessment as well. As an example, the Industrial Research Institute (IRI), whose 260 member companies invest over $55 billion annually in R&D, has shown intense interest in measuring research performance and effectiveness. The IRI has commissioned one of its internal panels to research the field and write a position paper on measuring and improving effectiveness of R&D on company performance.

When the above activities are integrated and placed into a mosaic, the inescapeable trend for the future becomes clear. The research sponsoring agencies will become more accountable to the Administration and Congress on the relationship between sponsored programs and strategic goals, and soon thereafter the research performers will become more accountable to the sponsoring agencies. In addition, the accountability of industrial research to the broader corporate goals will increase (as has been observed over the past decade), and improved methods of measuring research performance and productivity will be sought continually by industrial research organizations. It is therefore important that research managers and administrators in government, industry, and academia understand the assessment approaches which could be utilized to evaluate research quality and goal relevance, and that researchers gain an understanding of these evaluation approaches as well.

### *C. Underutilization of RIA*

Research, the pursuit and production of knowledge, has become a substantial investment in the U. S. and the rest of the developed world today. Depending on what is defined specifically as research in practice, public and private investment in research in the U. S. alone amounts to tens of billions of dollars per year. In 1990, for example, Federal support for basic and applied research approximated $22B, about 47 percent of total support for research in the U.S.[4] Typically, with investments of this magnitude, project selection and management are performed using the latest techniques available. Project payoff is estimated using the latest techniques and algorithms available. In addition, assessments of a large magnitude investment are done on a continuing basis, and there is a continual feedback loop to assure the investment will achieve its goals and targets.

While the methods used in the performance of research continually advance the state-of-the-art, the methods used for its identification and selection have changed little in decades. In evaluation and assessment of existing and completed research, not only have the methods in practice changed little with time, but the numbers of organizations which use any but the most rudimentary methods also remain a handful. While the scientific and social science literatures abound with advanced methodologies for identifying and selecting new research, managing existing research, and evaluating and assessing research retrospectively, the implementation of these methods by the research sponsoring community remains minimal.

### *D. Reasons for underutilization of RIA*

The reasons for reluctance to implement RIA vary. The rewards in research and research management go to new discoveries, not for quality assessments. Neither the costs nor time requirements of RIA are negligible, and have to be weighed against additional research which could be performed. More immediate organizational requirements are assigned higher priority than RIA. For example, an OTA assessment of the defense technology base states: "OSD [Office of the Secretary of Defense-RNK] personnel spend a large part of their time defending technology base programs or answering congressional mail, leaving little time available to evaluate technology base programs".[5]

The RIA outcomes are not always predictable or positive from a micro viewpoint, and 'pet' projects may be terminated after a rigorous evaluation. Any negative results from an RIA may provide executive or legislative branch overseers, or corporate management, ammunition for budget reductions. Finally, since there is very little experience with use of advanced evaluation techniques, there is insufficient evidence at present that use of advanced evaluation techniques will result in better payoff than use of rudimentary techniques. To many research managers and administrators, there is little to be gained from RIA, and a potential for loss.

*E. Benefits of increased utilization of RIA*

However, with the ascendency of Total Quality Management in many organizations, and with decreasing budgets and increased competitiveness at many levels, the motivation for a better understanding of the quantitative and qualitative measures of research impact has escalated in importance. Motivation to incorporate RIA into a permanent component of an organization's mode of operation, and determination to use the latest technological advances consistent with an organization's RIA requirement could have significant consequences at the organizational and national levels.

One major benefit would be to improve organizational efficiency. A properly executed RIA would target the people and the exogenous variables (management climate, funding conditions, infrastructure, etc.) necessary to increase research output relevant to the organization's goals. An RIA which increased communication among the researchers and potential research customers during the conduct of research would allow a smoother conversion of the products of research to technology through better integration of the users with the research performers.

Another major benefit would be to identify the diverse impacts of basic research. The impacts of basic research are pervasive throughout a technological society, but for the most part the impacts of basic research are indirect on technologies, systems, and end products. A major limitation of articulating the benefits of basic research has been the lack of data which could show the pathways and linkages through which the research impacts the intermediate or end products. A credible RIA of completed research would trace the dissemination of the research products through the many communication channels and would identify the multitude of near and long term research impacts (impact on other research fields, impact on technology, impact on systems, impact on education, etc.). Having this data would provide more substantive arguments for continuing to provide the necessary funds to those who control the allocation of research funds.

*F. Recent RIA studies*

One objective of the author's recent studies and the present paper is to identify many of the advanced and credible RIA approaches in use, or available today, and to enumerate both their strengths and weaknesses. Since research impact has many facets, its assessment must use as many methods and as many types of experts as required to address as many of these components as possible. Credible assessments will then weight the results of the different facet assessments relative to the different organizational goals, and arrive at conclusions optimal to the organization's interests.

Combinations of RIA approaches are recommended when performing a full assessment. While the readers schooled in systems reliability may question how the results from multiple imperfect approaches are improved as the number of approaches increase, experience has shown that a more acceptable product does result when different approaches are used. The effect appears to be additive rather than multiplicative. When different RIA approaches result in similar findings, the user will have confidence in the general theme of the results. When different approaches produce conflicting results, much value and understanding is gained by trying to understand the causes of the differences and trying to then resolve these differences.

Another objective of the recent studies and this paper is to show, somewhat indirectly, that while there is a significant gap between the RIA methods available in the literature and the RIA methods actually in use, there is also a substantial gap between the technologies becoming available from the research laboratories (such as information management and processing) and the technologies employed in the published methods. In the U. S., Federal support for developing the assessment methodologies which use the latest technologies has lagged other parts of the world. A cursory reading of the relevant literature shows that in the past two decades the U.S. efforts in this field have advanced at a very slow pace, and in many subfields *the*

*U. S. has been surpassed by other nations, notably those of Western Europe*. If it is assumed that improved RIA will lead to a more efficient allocation of research resources, then in the highly competitive research and technology based world which has evolved, the U. S. cannot afford to continue business as usual in its treatment of research and its impacts. It is hoped that this paper will help spur the Federal government, and private sources as well, to focus a concerted effort in advancing the techniques and implementation of RIA.

This paper is divided into three segments, which range from qualitative to quantitative approaches. The first segment deals with qualitative approaches to RIA. Foremost among these are variants on the common theme of peer review. While peer review (evaluation of research and its consequences by 'peers', or experts on the different facets of research and its impacts) is the method used most widely to evaluate research, it has its detractors, as will be shown in this paper. Because of cost and subjectivity, other methods to complement or replace peer review, and which are perhaps less costly and more objective, are being actively pursued.

The second segment deals with semi-quantitative approaches. These methods make little use of mathematical tools but attempt to draw on documented approaches and results wherever possible. They have limited credibility in the analytic community, since the selection of innovations to be analyzed tends to be arbitrary rather than mathematically rigorous, and they are viewed more as anecdotal approaches than serious technical approaches. Nevertheless, in practice, some of these approaches (namely, studies of accomplishments resulting from sponsored research programs, or studies of systems and the research products which were eventually converted and incorporated into those systems) are widely used by the research sponsoring organizations.

The third segment deals with the quantitative and fiscal approaches to RIA. These approaches make heavy use of mathematical and analytic tools, and utilize computer capabilities extensively. Probably the heaviest concentration of literature papers today are in this category. It should be noted that there are hybrid techniques which span more than one of the three categories. For example, a recent retrospective study of significant events in Cancer research[6] included a bibliometric component (citation and co-citation analyses).

## II. Research impact assessment methods

### *A. Background and overview*

There are some general principles, findings, and conclusions when the different methods described in this paper and their results are integrated and interpreted. First and foremost *is the role of motivation and associated incentives*. The research managers and administrators, and those with responsibility for higher level oversight, have to be convinced of the value of RIA to their organizations for the improved allocation of research resources. More important than any evaluation criteria selected is the dedication of an organization's management to the highest quality objective review, and the associated emplacement of rewards and incentives to encourage quality reviews. The team assigned responsibility to carry out RIA must be motivated to generate the highest quality product, not just 'answer the mail', as is done in many organizations today. This means selecting the best suite of methods available to accomplish organizational objectives, and selecting the most competent and objective individuals to participate in the RIA. The RIA managers must be motivated to examine the impact from as many perspectives as possible, to gain the most complete understanding. Finally, the objectives, importance, and benefits of RIA must be articulated and communicated to the researchers and research managers at the initiation of RIA, so that the reviewees will participate in the RIA as fully and as cooperatively as possible.

The total R&D process in an organization should be designed to include RIA as an *integral component*, not as an afterthought or an add-on. This will allow an orderly and continuous monitoring of the full research selection, review, and post-mortem analysis process, and insure that the best research consistent with the organization's goals is being funded. The evaluation methods selected should not be overly complex or require massive permanent staffs, and should offer minimum interference in the performance of the research.[7] Most managers regard applying overly elaborate and rigorous-seeming techniques to industrial R&D as inappropriate.[8] A reasonable fraction of the R&D budget should be allocated for RIA purposes, and advancement along a career path for RIA professionals should parallel that of the research performers.

An RIA should be conducted with maximum access to, and awareness of, information about research and technology development being pursued throughout the world. Access of the RIA to existing technology information would also be useful. This information will help determine whether the research being assessed is breaking new ground and, for high-tech organizations, whether the research being assessed is improving existing or developing technology.

Optimally, a database which contains this information would be available to those conducting an RIA. While subsets of this type of database exist, such as the Federal multiagency funded research programs database developed by the author, a comprehensive research and (developing and existing) technology database remains to be developed. Construction of such a database would require cooperation among Federal research sponsoring agencies and private organizations, at a minimum. As a starting point, Federal research sponsoring agencies should begin to coordinate requirements for such a database.

For organizations which sponsor substantial basic research, the RIA should be structured to identify impacts which occur *many decades after the research is performed*. The reasons for this are twofold. First, the impacts of basic research on organizational missions such as systems and operations can take decades before they are realized. Second, these organizational mission impacts will provide data for predictive models that relate research evaluation results to organizational mission impacts. Also, the *indirect impacts* of the research must receive a proper accounting. These indirect impacts contribute to an ever expanding pool of knowledge, and it is the *level of this pool which serves as the critical path to limiting the rate of advance of mission-oriented research*, and thereby technology and systems growth. While the determination of indirect impacts is complex and data intensive, it is absolutely necessary for a credible RIA.

The present paper addresses the predictive reliability of the RIA processes very briefly, mainly because there is little literature which provides the basis for predicting which research programs/proposals will have the desired downstream impact. For example, the relationship between a proposal's peer review score or a project's bibliometric rating and the downstream impact on an organization's mission is not addressed in published studies. One could raise the question, as many active researchers have, as to whether there is value to any of these assessment techniques, since their predictive value is unknown. The credibility and predictability of these assessment techniques are ripe topics for research. A long term tracking system for research product evolution would be required to gather the necessary data. The system would require agreement and coordination from a number of the larger Federal research sponsoring agencies, and maybe from industrial organizations as well. While such a system would not provide absolute answers, since tracking of the informal modes of knowledge communication would be almost impossible, it would provide a much better picture of research impact and its predictability than exists now. With the present state of information storage and processing capabilities, *research product evolution tracking is an idea whose time has come.*

### *B. Qualitative (peer review) methods*

#### *1. Background.*

Peer review of research represents evaluation by experts in the field, and is the method of choice in practice in the U. S.[9-15] Its objectives range from being an efficient resource allocation mechanism to a credible predictor of research impact.

#### *2. Requirements for high quality peer review.*

Many studies related to peer review have been reported in the literature, ranging from the mechanics of conducting a peer review, to examples of peer reviews, to detailed critiques of peer reviews and the process itself (e.g. Refs 9-12, 16-29). A non-standard peer review approach for concept comparisons is the Science Court. As in a legal procedure, it has well defined advocates, critics, a jury, etc. It was applied by the author to a review of alternate fusion concepts in 1977.[30] This procedure had substantial debate and surfacing of crucial issues, but it was time-consuming compared to a standard panel assessment.

While these reported studies present the process mechanics, the procedures followed, and the review results, the reader cannot ascertain the *quality* of the review and the results. In practice, procedure and process quality are mildly necessary, but nowhere sufficient, conditions for generating a high quality peer review. Many useful peer reviews have been conducted using a broad variety of processes, and while well documented modern processes (e.g., DOE[24]) may contribute to the efficiency of conducting a review, more than process is needed for high quality. There are many intangible factors that enter into a high quality review, and before examples of reviews are presented in the main body of this paper, some of the more important factors will be discussed.

The desirable characteristics of a peer review can be summarized as:[12]

- an effective resource allocation mechanism;
- an efficient resource allocator;
- a promoter of science accountability;
- a mechanism for policymakers to direct scientific effort;

- a rational process;
- a fair process;
- a valid and reliable measure of scientific performance.

High quality peer reviews require as a minimum the conditions summarized from *Ormala*:[27]

- The method, organization and criteria for an evaluation should be chosen and adjusted to the particular evaluation situation;
- Different levels of evaluation require different evaluation methods;
- Program and project goals are important considerations when an evaluation study is carried out;
- The basic motive behind an evaluation and the relationships between an evaluation and decision making should be openly communicated to all the parties involved;
- The aims of an evaluation should be explicitly formulated;
- The credibility of an evaluation should always be carefully established;
- The prerequisites for the effective utilization of evaluation results should be taken into consideration in evaluation design.

Assuming these considerations have been taken into account, *three of the most important intangible factors for a successful peer review are: Motivation, Competence, and Independence*. The review leader's motivation to conduct a technically credible review is the cornerstone of a successful review. The leader selects the reviewers, summarizes their comments, guides the questions and discussions in a panel review, and makes recommendations about whether the proposal should be funded. The quality of a review will never go beyond the competence of the reviewers. Two dimensions of competence which should be considered for a research review are the individual reviewer's technical competence for the subject area, and the competence of the review group as a body to cover the different facets of research issues (other research impacts, technology and mission considerations and impacts, infrastructure, political and social impacts). The quality of a review is limited by the biases and conflicts of the reviewers. The biases and conflicts of the reviewers selected should be known to the leader and to each other.

A broad range of reviewer expertise enhances the review results substantially. A key component of the process reported in *Kostoff*[26] was the use of mixed levels of reviewers on the panels to evaluate the different potential impacts of research. The panels included:

- bench-level researchers to address the impact of the proposed research on its field;
- broad research managers to address potential impact on allied research fields;
- technologists to address potential impact on technology and the potential of the research to transition to higher levels of development;
- systems specialists to address potential impact on systems and hardware;
- operational naval officers to address the potential impact on naval operations.

The presence of reviewers with different research target perspectives and levels of understanding on one panel provided a depth and breadth of comprehension of the different facets of the research impact that could not be achieved by segregating the science and utility components into separate panels and discussions.

Nearer-term research impacts typically play a more important role in the review outcome than longer-term impacts, but do not have quite the importance of team quality, research approach, or the research merit. A minimal set of review criteria should include team quality, research merit, research approach, productivity, and mission relevance.

The best features of different organizations' peer review practices can be combined into a heuristic protocol for the conduct of successful peer review research evaluations and impact assessments. The main aims of the protocol are to insure that the final assessment product has the highest intrinsic quality and that the assessment process and product are perceived as having the highest possible credibility. The protocol elements are:

*a. Peer review research evaluations.*

– The objectives of the assessment must be stated clearly and unambiguously at the initiation of the assessment by the highest levels of management, and the full support of top management must be given to the assessment. In turn, the objectives, importance, and urgency of the assessment must be articulated and communicated down the management hierarchy to the managers and performers whose research is to be assessed, and the cooperation of these reviewees must be enlisted at the earliest stages of the assessment;

– The final assessment product, the audience for the product, and the use to be made of the product by the audience should be considered carefully in the design of the assessment;

– One person should be assigned to manage the assessment at the earliest stage, and this person should be given full authority and responsibility for the assessment;

– The assessment manager should report to the highest organizational level possible in order to insure maximum independence from the research units being assessed;

– The reviewers should be selected to represent a wide variety of viewpoints, in order to address the many different facets of research and its impact.[26] These would include bench-level researchers to address the impact of the proposed research on the field itself; broad research managers to address potential impact on allied research fields; technologists to address potential impact on technology and the potential of the research to transition to higher levels of development; systems specialists to address potential impact on systems and hardware; and operational personnel to address the potential impact on downstream organizational operations. The reviewers should be independent of the research units being evaluated, and independent of the assessing organization where possible. The objectives of, and constraints on (if any), the assessment should be communicated to the reviewers at the initial contact;

– Maximum background material describing the research to be assessed, related research and technology development sponsored by external organizations, the organization structure, and other factors pertinent to the assessment, should be provided to the reviewers as early as possible before the review. This will allow the reviewers and presenters to use their time most productively during the review;

– Recommendations resulting from the assessment should be tracked to insure that they are considered and implemented, where appropriate. For research programs, planning, execution, and review are linked intimately. Feedback from the review outcomes to planning for the next cycle should be tracked to insure that the review/planning coupling is operable.

*b. Levels of organizational research evaluation.*

– Evaluations should be performed at three levels of resolution in the organization.

– The highest level would be an annual *corporate level review* of how the organization performs research. If the organization has a separate research unit, then the unit should be evaluated as an integrated whole. If research is vertically integrated with development, then the research should preferably be evaluated as part of a total organization R&D review. The charter of this highest level assessment would be to review, at the corporate level, general policy, organization, budget, and programs (e.g. Ref. 31). Total inputs and outputs, including integrated bibliometric indicators, would be examined. Overall research management processes would be examined, such as selection, execution, review, and technology transfer of research. The overall investment strategy would be evaluated, and would include different perspectives of the program, such as technical discipline, performer, and end use allocation. The integration of the research objectives with the larger organization objectives would be assessed. The evaluators would include, but not be limited to, representatives of the stakeholder, customer, and user community whose potential conflicts with the organization are minimal.

– The second level would be trienniel peer review of a discipline or management unit at the program level (e.g. Refs 15, 26), where a program is defined as an aggregation of work units (Principal Investigators). If the organization has a separate research unit, then the discipline should be evaluated as an integrated whole. In the nominal review, quality and relevance could be evaluated concurrently. If research is vertically integrated with development, then the research should preferably be evaluated as part of a total vertical structure R&D review. In the nominal vertical structure review, quality and relevance should preferably be evaluated separately. Thus, research evaluation must take into account how research is structured, integrated, and managed within an organization. Research quality criteria should include research merit, research approach, productivity, and team quality. Relevance criteria should include short term impact (transitions and/or utility), long term potential impact, and some estimate of the probability of success of attaining each type of impact. While the emphasis is on peer review, bibliometric and other type of indicators should be utilized to supplement the peer evaluation.

– The third level would be a minimum of trienniel peer review at the work unit (Principal Investigator) level (e.g. Ref. 24). Most of the program level issues described above are applicable and need not be repeated here.

– For each of these three levels of review, the following criteria and issues should be considered during the review as appropriate.

*c. Criteria for organization reviews.*

- Quality and uniqueness of the work;
- Scientific and technological opportunities in areas of likely organization mission importance;
- Need to establish a balance between revolutionary and evolutionary work;

- Position of the work relative to the forefront of other efforts;
- Responsiveness to present and future organization mission requirements;
- Possibilities of follow-on programs in higher R&D categories;
- Appropriateness of the efforts for organization vice other organizations;
- Other organization connection ( coordination) of the work.

*d. Questions to be asked of organization programs.*

- What is the investment strategy of the larger management unit? This would include the relative program priorities, the actual investment allocation to the different programs, and the rationale for the investment allocation. For each program being reviewed, what is the investment strategy for its thrust areas?
- What are we trying to do (in a systems concept)?
- Can specific advantage to the organization be identified if program is successful?
- How is the system done today and what are the limitations of the current practice?
- Would the work be supported if it were not already underway?
- Assuming success, what difference does it make to the user in a mission area content?
- What is the technical content of the program and how does it fit with other ongoing efforts in academia, industry, organization labs, other labs, etc.?
- What are the decision milestones of the program?
- How long will the program take; how much will the program cost; what are the mid-term and final objectives of the program?

In Europe, another development line has been to commission evaluation experts either to support panels or to conduct independent assessments which may involve surveys, in-depth interviews, case studies, etc.[32] *Barker*[16] describes how evaluation experts coming from two main communities (civil servants and academic policy researchers) interact in evaluation of R&D in the UK. The performance of evaluations, including the synthesis of evidence and the production of conclusions and recommendations, is done by professionals, as opposed to panels of eminent persons.

### 3. Problems with peer review

Peer review problems include:[11,12,33-35]

- Partiality of peers to impact the outcome for non-technical reasons;
- An 'Old Boy' network to protect established fields;
- A 'Halo' effect for higher likelihood of funding for more visible scientists/ departments/ institutions;
- Reviewers differ in criteria to assess and interpret;
- The peer review process assumes agreement about what good research is, and what are promising opportunities.

These potential problems should be considered during the process of selecting research impact assessment approaches.

Another problem with peer review is cost. The true total costs of peer review can be considerable but tend to be ignored or understated in most reported cases. For serious panel-type peer reviews, where sufficient expertise is represented on the panels, *total real costs will dominate direct costs* by as much as an order of magnitude or more.[14] The major contributor to total costs for either type of review is the time of all the players involved in executing the review. With high quality performers and reviewers, time costs are high, and the total review costs can be a non-negligible fraction of total program costs, especially for programs that are people intensive rather than hardware intensive.

The issue of peer review predictability affects the credibility of technological forecasting directly. A few studies have been done relating reviewers' scores on component evaluation criteria to proposal or project review outcomes. Some studies have been done in which reviewers' ratings of research papers have been compared to the numbers of citations received by these papers over time.[36,37] Correlations between reviewers' estimates of manuscript quality and impact and the number of citations received by the paper over time were relatively low. The author is not aware of reported studies, singly or in tandem, that have related peer review scores/rankings of proposals to *downstream impacts* of the research on technology, systems, and operations. This type of study would require an elaborate data tracking system over lengthy time periods which does not exist today. Thus, the value of peer review as a predictive tool for assessing the impact of research on an organization's mission (other than research for its own sake) rests on faith more than on hard documented evidence.

*4. Peer review conclusions.*

Peer review is the most widely used and generally credible method used to assess the impact of research. Much of the criticism of peer review has arisen from misunderstandings of its accuracy resolution as a measuring instrument. While a peer review can gain consensus on the projects and proposals that are either outstanding or poor, there will be differences of opinion on the projects and proposals that cover the much wider middle range. For projects or proposals in this middle range, their fate is somewhat more sensitive to the reviewers selected. If a key purpose of a peer review is to insure that the outstanding projects and proposals are funded or continued, and the poor projects are either terminated or modified strongly, then the capabilities of the peer review instrument are well matched to its requirements.

However, the value of peer review as a predictive tool for assessing the impact of research on an *organization's mission* (other than research for its own sake) rests on faith more than on hard documented evidence. Also, for serious panel-type peer reviews or mail-type peer reviews, where sufficient expertise is represented on the panels, *total real costs will dominate direct costs*. The major contributor to total costs is the time of all the players involved in executing the review. With high quality performers and reviewers, time costs are high, and the total review costs can be a non-negligible fraction of total program costs, especially for programs that are people intensive rather than hardware intensive.

Most methods used in practice include criteria which address the impact of research on its own and allied fields, as well as on the mission of the sponsoring organization. Nearer-term research impacts typically play a more important role in the review outcome than longer-term impacts, but do not have quite the importance of team quality, research approach, or the research merit. A minimal set of review criteria should include team quality, research merit, research approach, research productivity, and a criterion related to longer-term relevance to the organization's mission. More important than the criteria is the dedication of an organization's management to the highest quality objective review, and the associated emplacement of rewards and incentives to encourage quality reviews.

### *C. Retrospective methods*

#### *1. Background and overview*

In the evaluation of research performance and impact, a spectrum of approaches may be considered. At one end of the spectrum are the subjective, essentially non-quantitative approaches, of which peer review is the prototype.[12] At the other end of the spectrum are the mainly quantitative approaches, such as evaluative bibliometrics and cost-benefit.[38,39] In between are retrospective or case study approaches.[40,41]

These retrospective methods make little use of mathematical tools, but draw on documented approaches and results wherever possible. In practice, there are two major reasons that research sponsoring organizations perform retrospective studies of research. Positive research impact on the organization's mission provides evidence to the stakeholders that there is benefit in continuing sponsorship of research. Also, if the study is sufficiently comprehensive, the environmental parameters which helped the research succeed can be identified, and these lessons can be used to improve future research.

There are two major variants of retrospective studies. One type starts with a successful technology or system and works backwards to identify the critical R&D events which led to the end product. The other type starts with initial research grants and traces evolution forward to identify impacts. The tracing backwards approach is favored for two reasons: (1) the data is easier to obtain, since forward tracking is essentially non-existent for evolving research; and (2) the sponsors have little interest in examining research that may have gone nowhere.

While methods for performing retrospective and case studies may differ within and across industry and government,[41] especially concerning the research question, case selection, and analytic framework, the fundamental evaluation problems encountered are pervasive across these different methods. In the remainder of this summary, a few of the more widely known case studies will be reviewed, and the key pervasive problems and findings will be identified. These retrospective studies include Project Hindsight, Project TRACES and its follow-on studies, and Accomplishments of Department of Energy (DOE) Office of Health and Environmental Research (OHER) and of the Advanced Research Projects Agency (ARPA).

*2. Specific retrospective studies.*

*i. Project Hindsight.*

Project Hindsight was a retrospective study performed by the Defense Department in the mid-1960s to identify those management factors important in assuring that research and technology programs are productive and that program results are used.[42] The evolution of the new technology represented in each of the 20 weapons systems selected was traced back in post-WW2 time to critical points called "Research or Exploratory Development (RXD) Events".

*ii. Original TRACES study.*

In 1967, The National Science Foundation (NSF) instituted a study[43] to trace retrospectively key events which had led to a number of major technological innovations. One goal was to provide more specific information on the role of the various mechanisms, institutions, and types of R&D activity required for successful technological innovation. Similar to Project Hindsight, key 'events' in the R&D history of each innovation selected were identified, and their characteristics were examined.

*iii. Follow-on TRACES study.*

In a follow-on study to TRACES, the NSF sponsored Battelle-Columbus Laboratories to perform a case study examination of the process and mechanism of technological innovation.[44] For each of the ten innovations studied, the significant events (important activity in the history of an innovation) and decisive events (a significant event which provides a major and essential impetus to the innovation) which contributed to the innovation were identified. The influence of various exogenous factors on the decisive events was determined, and several important characteristics of the innovative process as a whole were obtained. The following important exogenous factors for producing significant innovations were identified:

- The technical entrepreneur (a major driving force in the innovative process);
- Early recognition of the need;
- Government funding (more generally, availability of financial support, from whatever source);

- The occurrence of an unplanned confluence of technology (confluence of technology occurred for some innovations as a result of deliberate planning, rather than by accident);
- Most of the innovations originated outside the organization that developed them;
- Additional supporting inventions were required during the development effort for all the innovations studied to arrive at a product with consumer acceptance.

*While the technical entrepreneur is viewed as extremely important to the innovative process, it does not appear (to the author) to be the critical path factor. Examination of the historiographic tracings which display the significant events chronologically for each of the innovations shows that an advanced pool of knowledge must be developed in many fields before synthesis leading to an innovation can occur. The entrepreneur can be viewed as an individual or group with the ability to assimilate this diverse information and exploit it for further development. However, once this pool of knowledge exists, there are many persons or groups with capability to exploit the information, and thus the real critical path to the innovation is more likely the knowledge pool than any particular entrepreneur. The entrepreneurs listed in the study undoubtedly accelerated the introduction of the innovation, but they were at all times paced by the developmental level of the knowledge pool.*

*iv. Recent TRACES study.*

In a modern version of the TRACES study, the National Cancer Institute initiated an assessment[6] to determine whether there were certain research settings or support mechanisms which were more effective in bringing about important advances in cancer research. The approach taken was analogous in concept to the initial TRACES study, with the addition of citation analyses to provide an independent measure of the impact of the Trace papers (papers associated with each key 'event'), and by adding control sets of papers.

*v. DARPA Accomplishments study.*

The Institute for Defense Analysis produced a document[45] describing the accomplishments of the Defense Advanced Research Projects Agency (DARPA-now renamed ARPA). Of the hundreds of projects and programs funded by DARPA over its then (1988) 30 year lifetime, 49 were selected and studied in detail, and conditions for success were identified.

The qualities of DARPA-supported programs and projects that contributed to success can be summarized:

- A need existed for what the output could do;
- There was a strong commitment by individuals to a concept;
- Bright and imaginative individuals were given the opportunity to pursue ideas with minimal bureaucratic encumbrance;
- There was an ongoing stream of technical developments and evolution;
- DARPA management gave strong, top-level management support;
- There was explicit effort, taken early, to improve acceptance by the user community.

*vi. DOE OHER Accomplishments book.*

The approach taken by DOE was to describe the 40-year history of OHER,[46,47] and present selected accomplishments in different research areas from different points in time. This technique allowed impacts and benefits of the research to be tracked through time, and in some cases to be quantified as well.

*3. Retrospective studies conclusions.*

Hindsight, TRACES, and, to some degree, the OHER and DARPA accomplishments books had some similar themes. All these methods used a historiographic approach, looked for significant research or development events in the metamorphosis of research programs in their evolution to products, and attempted to convince the reader that: (1) the significant research and exploratory development events in the development of the product or process were the ones identified; (2) typically, the organization sponsoring the study was responsible for some of the (critical) significant events; (3) the final product or process to which these events contributed was important; and (4) while the costs of the research and development were not quantified, and the benefits (typically) were not quantified, the research and development were worth the cost.

Six critical conditions for innovation were identified through analysis of these retrospective studies. The most important condition appears to be *the existence of a broad pool of knowledge which minimizes critical path obstacles* and can be exploited for development purposes. This condition is followed in importance by a *technical entreprenuer who sees the technical opportunity and recognizes the need for innovation*, and who is willing to champion the concept for long time periods, if necessary. Also valuable are *strong financial and management support* coupled with *many continuing inventions in different areas* to support the innovation.

As the historiographic analyses (Hindsight/ TRACES) of a technology or system have shown, if the time interval in which the antecedent critical events occur is arbitrarily truncated, as in the two-decade time interval Hindsight case, the impacts of basic research on the technology or system will not be given adequate recognition. The number of mission oriented research events peaks about a decade before the technology innovation. However, the number of non-mission oriented research events peaks about three decades before the technology innovation, and eight, nine, or more decades may be necessary in some cases to recognize the original critical antecedent events. Over a long time interval, the majority of key R&D events tend to be non-mission oriented. Thus, future studies of this type should allow time intervals of many decades to insure that critical non-mission oriented research events are captured.

Even in those cases when an adequate time interval was used, and critical non-mission oriented events were identified, the cumulative indirect impacts of basic research were not accounted for by any of the retrospective approaches published or in use today. A recent study[48] which examined impacts of research on other research and technology through direct and indirect paths using a network approach showed that the indirect impacts of fundamental research can be very large in a cumulative sense. Future retrospective studies would be more credible if they devote more effort to identifying indirect impacts of research. While indirect impacts of research are much more difficult to identify than direct impacts, and the data gathering effort is much larger and more complex, neglect of indirect impacts reduces appreciation of the value of basic research significantly. Use of some of the advanced computer-based technologies available today, such as the network approach referenced above or citation analysis,[6] could identify many of the pathways of the indirect impacts of research.

A detailed reading of those studies which attempted to incorporate economic quantification showed the difficulties of trying to identify, assign, and quantify costs and benefits of basic research, especially at a project/investigator level. As TRACES and other similar studies have shown, the chain of events leading to an innovation is long and broad. Many researchers over many years have been involved in the chain, and many funding agencies, some simultaneously with the same researchers, may have been involved. The allocation of costs and benefits under such circumstances is a very difficult and highly arbitrary process. The allocation problem is reduced, but not eliminated, when the analysis is applied at the macro level (integrating across individual researchers, organizations, etc.).

One goal of all the studies presented was to identify the products of research and some of their impacts. The Hindsight, TRACES, and ARPA studies tried to identify factors which influenced the productivity and impact of research. The following conclusions about the role and impact of basic research were reached:

- The majority of basic research events which directly impacted technologies or systems were non-mission oriented and occurred many decades before the technology or system emerged;
- The cumulative indirect impacts of basic research were not accounted for by any of the retrospective approaches published;
- An advanced pool of knowledge must be developed in many fields before synthesis leading to an innovation can occur;
- Allocation of benefits among researchers, organizations, and funding agencies to determine economic returns from basic research is very difficult and arbitrary, especially at the micro level.

While these approaches do provide interesting information and insight into the transition process from research to development to products, processes, or systems, the *arbitrary selectivity and anecdotal nature of many of the results render any conclusions as to cost-effectiveness or generalizability suspect*. Supplementary analyses using other approaches are required for further justification of the value of the R&D.

### *D. Quantitative methods*

#### *1. Background.*

Quantitative approaches to research assessment focus on the numerics associated with the performance and outcomes of research. The main approaches used are bibliometrics and econometrics such as cost-benefit and production function analysis. This summary focuses on these three main approaches, briefly describes the bibliometrics-related family of approaches known as co-occurrence phenomena, briefly describes a network modeling approach to quantifying research impacts, and ends with an expert systems approach for supporting research assessment.

*2. Bibliometrics.*

*a. Foundations.*

Bibliometrics, especially evaluative bibliometrics, uses counts of publications, patents, citations and other potentially informative items to develop science and technology performance indicators. The choice of important bibliometric indicators to use for research performance measurement may not be straightforward. A 1993 study surveyed about 4,000 researchers to identify appropriate bibliometric indicators for their particular disciplines.[49] The respondents were grouped in major discipline categories across a broad spectrum of research areas. While the major discipline categories agreed on the importance of publications in refereed journals as a performance indicator, there was not agreement about the relative values of the remaining 19 indicators provided to the respondents. For the respondents in total, the important performance indicators were:

- Publications (publication of research results in refereed journals);
- Peer Reviewed Books (research results published as commercial books reviewed by peers);
- Keynote Addresses (invitations to deliver keynote addresses, or present refereed papers and other refereed presentations at major conferences related to one's profession);
- Conference Proceedings (publication of research results in refereed conference proceedings);
- Citation Impact (publication of research results in journals weighted by citation impact);
- Chapters in Books (research results published as chapters in commercial books reviewed by peers);
- Competitive Grants (ability to attract competitive, peer reviewed grants from the ARC, NH&MRC, rural R&D corporations and similar government agencies).

These bibliometric indicators can be used as part of an analytical process to measure scientific and technological accomplishment. Because of the volume of documented scientific and technological accomplishments being produced (5,000 scientific papers published in refereed scientific journals every working day worldwide; 1,000 new patent documents issued every working day worldwide), use of computerized analyses incorporating quantitative indicators is necessary to understand the implications of this technical output.[38]

*Narin* states three axioms that underlie the utilization and validity of bibliometric analysis. The first axiom is *activity measurement*: that counts of patents and papers provide valid indicators of R&D activity in the subject areas of those patents or papers, and at the institution from which they originate. The second axiom is *impact measurement*: that the number of times those patents or papers are cited in subsequent patents or papers provides valid indicators of the impact or importance of the cited patents and papers. However, there could be weightings applied to the raw count data, depending on the perceived importance of the journals containing the citing papers. Also, the impacts would be on allied research fields or technologies, not necessarily long-term impacts on the originating organization's mission. The third axiom is *linkage measurement*: that the citations from papers to papers, from patents to patents and from patents to papers provide indicators of intellectual linkages between the organizations which are producing the patents and papers, and knowledge linkage between their subject areas.[38]

Use of bibliometrics can be categorized into four levels of aggregation:[38]

- Policy (evaluation of national or regional technical performance);
- Strategy (evaluation of the scientific performance of universities or the technological performance of companies);
- Tactics (tracing and tracking R&D activity in specific scientific and technological areas or problems);
- Conventional (identifying specific activities and specific people engaged in research and development).

Policy questions deal with the analysis of very large numbers of papers and patents, often hundreds of thousands at a time, to characterize the scientific and technological output of nations and regions. Strategic analyses tend to deal with thousands to tens of thousands of papers or patents at a time, numbers that characterize the publication or patent output of universities and companies. Tactical analyses tend to deal with hundreds to thousands of papers or patents, and deal typically with activity within a specific subject area. Finally, conventional information retrieval tends to deal with identifying individual papers, patents, inventors and clusters of interest to an individual scientist or engineer or research manager working on a specific research project.

The first, and major, step in the performance of a high quality bibliometric analysis in any of the above four levels of aggregation is acceptance by the potential user of the above three axioms to validate the credibility of the bibliometric approach. Once this hurdle has been passed, the second step is to select the highest quality and reliability raw indicator products (data and databases) and apply analyses of the highest statistical precision and accuracy to these indicators.[50-52] The third step, which in many cases will determine the utility of the results, is the interpretation and visual display of the results. The results of the most stringent analyses will be relatively worthless if they are not displayed in a concise and lucid form. Indicators can be arranged in one or more dimensions. Emphasis has always been laid on the necessity of multidimensional thinking while analyzing scientometric indicators. Scientific research is a multifaceted human activity, and overemphasizing any of its aspects (publication productivity, citation influence, technological applicability, etc.) may lead to serious distortions in its assessment. While each scientometric indicator represents a single component of a multidimensional manifold which itself is just one element in assessing a complex system, presentations in one or several dimensions may equally prove useful.[52]

The most direct way of presenting scientometric indicators is in one dimensional ranked lists. While simplistic, this approach reflects the paramount competitiveness of the scientific enterprise. Linear rankings are most attractive for presentation to the larger non-specialist audience (see Ref. 52).

Two dimensional displays can include relational charts or scatter plots for correlations. In two dimensional relational charts,[53,54] pairs of indicators (observed vs. expected citation rates or attractivity vs. activity indices)are displayed in a planar orthogonal coordinate system. Emphasis is shifted from ranking to the formation of groups or 'clusters' and other characteristic relations among various indicators.

An obvious deficiency of the relational charts is the lack of any indication of the size of the sets of publications underlying the points of the diagram. By adding the third dimension of publication size, this objection can be overcome. The basic idea of 'landscaping' national scientific performances is to represent the size by the 'mass' of a mountain-like formation. If two or more countries have similar citation characteristics, the peaks representing them may get superimposed forming chains, massifs, and other surface formations. An example is presented in *Braun*.[55]

There seems to be a natural limit of graphical presentation at three dimensions. There are techniques, however, to overcome this apparent restriction. A rather original method of representing multivariate data was proposed by Herman Chernoff: "Each point in k-dimensional space, $k \leq 18$, is represented by a cartoon face whose features, such as length of nose and curvature of mouth correspond components of the point. Thus every multivariate observation is visualized as a computer drawn face. This presentation makes it easy for the human mind to grasp many of the essential regularities and irregularities present in the data."

*Braun*[52] shows a face pattern with 18 facial features applicable in representing multidimensional data. *Schubert*[56] contains a four-dimensional example of applying Chernoff-faces in scientometrics: uncitedness, citation rate per cited paper, mean expected citation rate and relative citation rate are represented by the shape of face, size of eyes, length of nose and curvature and length of mouth, respectively.

*b. Problems with bibliometrics.*

Problems with publication and citation counts include:[28,34,57,58]

(1) *Publication counts*:
- Indicates quantity of output, not quality;
- Non-journal methods of communication ignored;
- Publication practices vary across fields, journals, employing institutions;
- Choice of a suitable, inclusive database is problematical;
- Undesirable publishing practices (artificially inflated numbers of co-authors, artificially shorter papers) increasing.

(2) *Citations*:
- Intellectual link between citing source and reference article may not always exist;
- Incorrect work may be highly cited;
- Methodological papers among most highly cited;
- Self-citation may artificially inflate citation rates;
- Citations lost in automated searches due to spelling differences and inconsistencies;
- *Science Citation Index* (SCI) changes over time;
- SCI biased in favor of English language journals;
- Same problems as publication counts.

There are few Federally-supported bibliometric studies reported in the literature. In addition to the above problems, another reason for limited Federal use can be inferred from *Narin*,[59] where studies on the publication and citation distribution functions for individuals are reviewed. The conclusion drawn, from studies such as those of Lotka, Shockley, De Solla Price, and Cole and Cole, is that *very few of the active researchers are producing the heavily cited papers*. How motivated are funding agencies to report these hyperbolic productivity distributions for different programs in the open literature, especially since many questions exist as to the accuracy and completeness of the bibliometric indicators? This conclusion raises the further question of the role actually played by the less productive researchers (as measured by publication and citation counts): is the productivity of the elite somehow dependent on the output of the less influential, or is the role of the less productive members that of maintaining the stability of the research infrastructure and educating future generations of researchers?

*c. Bibliometric studies.*

Macroscale bibliometric studies characterize science activity at the national (e.g. Refs 50, 60), international, and discipline level. The biennial *Science and Engineering Indicators* report[61] tabulates data on characteristics of personnel in science, funds spent, publications and citations by country and field, and many other bibliometric indicators. Another study at the national level was aimed at evaluating the comparative international standing of British science.[62] Using publication counts and citation counts, the authors evaluated scientific output of different countries by technical discipline as a function of time. Much more understanding is required as to which indicators are appropriate and how they should impact allocation decisions.

There have been numerous microscale bibliometric studies reported in the literature (e.g. Refs 63-72). The NIH bibliometric-based evaluations[28] included the effectiveness of various research support mechanisms and training programs, the publication performance of the different institutes, the responsiveness of the research programs to their congressional mandate, and the comparative productivity of NIH-sponsored research and similar international programs.

Two papers[6,73] described determination of whether significant relationships existed among major cancer research events, funding mechanisms, and performer locations; compared the quality of research supported by large grants and small grants from the National Institute of Dental Research; evaluated patterns of publication of the NIH intramural programs as a measure of the research performance of NIH; and evaluated quality of research as a function of size of the extramural funding institution. Most of the NIH studies focused on aggregated comparison studies (large grants vs small, large schools vs small schools, domestic vs foreign, etc).

Patent citation analysis has the potential to provide insight to the conversion of science to technology.[74-80] Much of the Federal government support of the development of patent citation analysis was by the NSF (e.g. Refs 81, 82). Some recent studies have focused on utilization of patent citation analysis for corporate intelligence and planning purposes (e.g. Ref. 83). Some of the data presented verify further Lotka's Productivity Law, where relatively few people in a laboratory are producing large numbers of patents. In the example presented in Ref. 83, patents of the most productive inventor are highly cited, further demonstrating his importance. Narin concludes that highly productive research labs are built around a small number of highly productive, key individuals.

Despite its limitations, bibliometrics may have utility in providing insight into research product dissemination. For laboratories, these studies include:

- Examine distribution of disciplines in co-authored papers, to see whether the multidisciplinary strengths of the lab are being utilized fully;
- Examine distribution of organizations in co-authored papers, to determine the extent of lab collaboration with universities/ industry/ other labs and countries;
- Examine nature (basic/ applied) of citing journals and other media (patents), to ascertain whether lab's products are reaching the intended customer(s);
- Determine whether the lab has its share of high impact (heavily cited) papers and patents, viewed by some analysts as a requirement for technical leadership;
- Determine which countries are citing the lab's papers and patents, to see whether there is foreign exploitation of technology and in which disciplines;
- Identify papers and patents cited by the lab's papers and patents, to ascertain degree of lab's exploitation of foreign and other domestic technology.

A recent comparative bibliometric analysis of 53 laboratories[84] clustered the labs into six types (Regulation and Control, Project Management, Science Frontier, Service, Devices, Survey), and stated that "comparisons of scientific impacts should be made only with laboratories that are comparable in their primary task and research outputs". The report concluded further that:

- Bibliometric indicators and scientific publications are not the only outputs that should be measured, but the other types of outputs differ for different labs;
- Bibliometric indicators are not equally valid across different types of laboratories;
- Bibliometric indicators are less useful for the evaluation of research laboratories involved in closed publication markets.

*Potential Normalization Approaches*

A major problem with bibliometrics is comparisons of outputs of different performers (or performing organizations) who may also work in different disciplines. Three types of normalization solutions to allow cross-organization or cross-discipline comparisons are proposed by *Schubert*.[85]

1. *The Publishing Journal as Reference Standard*

By relating the number of citations received by a paper (or the average citation rate of a subset of papers published in the same journal – the *Mean Observed Citation Rate*) to the average citation rate of all papers in the journal (the *Mean Expected Citation Rate*) the *Relative Citation Rate* will be obtained. This indicator shows the relative standing of the paper (or set of papers) in question among its close companions: it value is higher\lower than unity as the sample is more\less cited than the average.

2. *The Set of Related Records as Reference Standard*

"Bibliographic Coupling" uses the number of references a given pair of documents have in common to measure the similarity of their subject matter. Comparing a set of papers that are "similar" in this sense to a given article of the same age will yield an ideal reference standard for citation assessments.

3. *The Set of Cited Journals as Reference Standard*

A promising method is based on the *journal in the reference lists of the articles of the journal in question*. These journals are selected by the most reliable persons, the authors of the journal as references (in both senses of the word) and therefore, can justly be regarded as standards of the expected citation rate.

*3. Co-occurrence phenomena.*

One class of computer-based analytic techniques which tends to focus more on macroscale impacts of research exploits the use of co-occurrence phenomena. In co-occurrence analysis, *phenomena that occur together frequently in some domain are assumed to be related, and the strength of that relationship is assumed to be related to the co-occurrence frequency*. Networks of these co-occurring phenomena are constructed, and then maps of evolving scientific fields are generated using the link-node values of the networks. Using these maps of science structure and evolution, the research policy analyst can develop a deeper understanding of the interrelationships among the different research fields and the impacts of external intervention, and can recommend new directions for more desirable research portfolios. These techniques are discussed in more detail in references.[86-89] The Tijssen paper contains an excellent exposition on mapping techniques for displaying the structure of related science and technology fields.

In particular, co-citation analysis has been applied to scientific fields, and co-citation clusters have been mapped to represent research-front specialties.[89] Co-word has been utilized to map the evolution of science under European (mainly French) government support, and has the potential to supplement other research impact evaluation approaches. Co-nomination, in its different incarnations, has been used to construct social networks of researchers and has the potential, if expanded to include research and technology impacts in the network link values, for evaluating direct and indirect impacts of research.[90] Co-classification is based on co-occurrences of classification codes in patents, and is used to construct maps of technology clusters.[91]

*4. Cost-benefit/economic analyses.*

A comprehensive survey examined the application of economic measures to the return on research and development as an investment in individual industries and at the national level.[28] This document concluded that while econometric methods have been useful for tracking private R&D investment within industries, the methods failed to produce consistent and useful results when applied to Federal R&D support. A more recent analysis focused on economic/ cost-benefit approaches used for research evaluation.[92] The methods involve computing impacts using market information, monetizing the impacts, then comparing the value of the impacts with the cost of research. Principal measures described include surplus measures and productivity measures. With known benefit and cost time streams, internal rates of return to R&D investments are then computed. The paper notes both the standard technical difficulties with these approaches and the political and organizational difficulties in implementing them.

*a. Cost-benefit.*

Cost-benefit analyses are a family of related techniques which include Cost-Benefit, Net Present Value, and Rate-of-Return.[92-94] These approaches tend to be more widely used in industry than government. For one, or many, projects, the basic approach is similar. A starting point in time for the research is defined. The time stream of costs for product development is estimated, and the time stream of benefits from the product is estimated. Using the time value of money, the costs and benefits are discounted to the origin of time, and the net benefits are compared with the net costs. The main differences in the approaches to cost benefit analyses are in the sophistication of the methods used to estimate the cost and benefit streams, and the time value of money.

Cost-benefit analyses have limited accuracy when applied to basic research because *of the quality of both the cost and benefit data due to the large uncertainties characteristic of the research process, as well as selection of a credible origin of time for the discounting computations*. As an illustrative example, a deterministic cost-benefit analysis was performed by the author on a fusion reactor variant.[95] Its *real problem, which pervades and limits any attempt to perform a cost-benefit analysis on a concept in the basic research stage, was the inherent uncertainty of controlling the fusion process. This translated to the inability to predict the probabilities of success and time and cost schedules for overcoming fundamental plasma research problems (e.g., plasma stabilities and confinement times); no credible methods were available*. Thus, the main value of the cost-benefit approach was to show that the potential existed for positive payoff from the hybrid reactor development, that there was a credible region in parameter space in which controlled fusion development could prove cost effective; what was missing was the likelihood of achieving that payoff.

A 1991 marginal cost-benefit study weighed the costs of academic research against the benefits realized from the earlier introduction of innovative products and processes due to the academic research.[39] The study used survey data to show a very high social rate of return resulting from academic research. While the method is innovative, future applications using more objective data sources would provide higher confidence in the computed rates of return.

*b. Production function.*

Production function approaches to evaluating research returns invoke economic theory-based assumptions relating outputs to inputs to generate an estimatable model. One only needs time series data on output, capital, labor, and research expenditures to estimate empirically the marginal contribution of research to value added. However, the relationship of research to value added is non-linear and indirect. Variables such as other inputs to technology and production and marketing functions complicate the research/ value added relationship.

Much of the major recent economic work relating economic growth/ productivity increases to R&D spending has been performed by three economists.[39,96-100] *Mansfield*'s earlier study typifies the strengths and weaknesses of the production function approach. This study[96] attempted to determine whether an industry's or firm's rate of productivity change was related to the amount of basic research it performed. *Mansfield* developed a production function which disaggregated basic and applied research, then regressed rate of productivity increase with many different variables. The regressions showed a strong relationship between the amount of basic research carried out by an industry and the industry's rate of productivity increase during 1948-1966.

The study exemplifies the problem inherent in multiple regression analyses: that of determining cause and effect from what is essentially correlation. As *Mansfield* points out, "It is possible that industries and firms with high rates of productivity growth tend to spend relatively large amounts on basic research, but that their high rates of productivity growth are not due to these expenditures".[96] Nor does Mansfield's model specify the path(s) by which R&D investment supposedly leads to productivity improvements.

A production function approach to cost-efficiency of basic research essentially used a regression analysis between outputs and inputs.[101-102] For proposals, the method involved regressing output variables (citations per dollar, graduate students per dollar) against input variables (e.g., quality of the investigator's department, quality of the investigator, etc.). The results gave some idea of the importance of the input variables, alone or in combination, on the output variables. One obvious potential application would be prediction of proposals likely to have high productivity based on prior (input) knowledge. *Much, however, remains to be done in identifying the appropriate output measures, the appropriate input measures, and the nature of the interactions among these measures for different disciplines.*

*5. Network modeling for direct/indirect impacts.*

A network based modeling approach was devised which would allow estimation of the direct and indirect impacts of a research program or collection of research programs. The research program impacts would be multi-faceted, including impacts on advancing its own field, on advancing allied fields, on advancing technology, on supporting operations and mission requirements, etc. A major feature of the model is *inclusion of feedback from the higher development categories (e.g., exploratory development, advanced development) on the advancement of research.*

The model and a subsequent pilot study related to Navy R&D have been described in detail.[48] In summary, a network was constructed in which each node represented an area of research or development. The values of the links connecting each node pair represented the impact of results from the first node area on the second node area. The total impact of an area of research on other research or development was obtained by integrating over all paths from the research node to the node(s) of interest.

*6. Expert networks.*

Research Impact Assessment is, at its essence, a diagnostic process with many diagnostic tools. In other fields of endeavor, such as Medicine and Machinery Repair, expert systems are increasingly being used as diagnostic tools or as support to diagnostic processes. Recently, there have been efforts to develop expert system approaches combined with artificial neural networks (expert networks) for use in R&D management, including RIA.[103-105] A brief summary of these efforts follows.

The product of these efforts is Research-Management Expert Network (*R-MEN*) which is characterized by two complementary tools: Organizational/Professional Development and Expert Network. The latter technology is comprised of an expert system (left side brain) and an artificial neural network (right side brain). Given a set of research, and research management policies and strategies, *R-MEN* learns concepts that hierarchically organize those policies and strategies and use them in classifying/triaging research proposals.

The framework of Research-Management Expert Network (*R-MEN*) consists of a knowledge base and a data base. Feeding into the knowledge base are four modules: a policy/ strategy impartation module and a proposal data acquisition module, both of which receive input from the O/PD process; and a research impact calculation module and a proposal review module. The knowledge base then feeds into the data base through five modules: a project selection module, resources allocation module, project evaluation and control module, investigator evaluation module, and organization evaluation module.

*R-MEN* is implemented in three phases. Phase 1 includes the development of the strategic plan, which defines and communicates longer-term research directions, and the development of the operating plan, which specifically identifies the projects that will implement the strategic plan taking into consideration the goals, quantifiable objectives and development of the individual investigator and the organization.

Phase 2 represents the necessary education, and management support needed to prepare the staff to participate in such an "Action Research" effort. This phase identifies and utilizes the critical components required to develop an environment that facilitates participative research management activities. A significant activity occurring during this phase is daily verification of individual scheduled training and development. If an individual has no recorded training and/or development within a preset period, the system will generate and send a report through E-mail directly to the office of the director for R&D. The system will be able to look at a training and/or development description(s) and compare it/them with the background of the individual to determine if the training and/or development is/are suitable for that individual.

Phase 3 represents a means by which participative methods can be put into operation in developing productivity tracking systems. Significant activities occurring during this phase include project evaluation and control. This entails periodic monitoring of project milestones for applied research, and research objectives for the more basic research. If a project has no recorded fulfillment of a milestone within a preset period, the system will generate and send a report through E-mail directly to the office of the director for R&D.

If *R-MEN* is initially used concurrently with present research review processes, it will serve as a supplement in the form of a guide to data generation, acquisition and processing, and a validity check. With appropriate implementation and maintenance, this knowledge technology, which utilizes demonstrated and proven approaches, methods, procedures and techniques in an innovative and unique way, could lead to the following benefits:

- Provide a means for effective, policy- and strategy-oriented management through outcomes-management.
- Improve management quality, reduce operation costs, and increase productivity and public trust.
- Foster impact evaluation to document Federally funded program and management effectiveness.
- Provide short-term (three-year) program progress tracking and long-term (ten-year) result(s) impact tracking.
- Shield administrators, managers, and other policy-makers from the complexity of the mathematics of the inference machine.
- Permit the evaluation of a range of alternatives.
- Permit handling large amounts of data.

- Permit policy-makers to have a better understanding of existing technical attributes of and capabilities for potential projects.
- Facilitate choice of strategy compatible with agency structure and processes, and with the policy or the nature of decision making for activities scheduling and control.

*7. Quantitative methods conclusions.*

Bibliometric methods are valuable in quantifying the output of research. Because they do not address quality, and their numeric outputs are subject to multiple interpretations, they are not self-contained assessment methods. They are a valuable supplement to the subjective interpretative methods such as peer review.

Economic approaches have limited value when applied to assessing the potential of fundamental research, because of the uncertain nature of the data. Their validity increases as the research becomes more applied, and cost and benefit streams can be estimated more accurately.

As databases become more extensive, and computer power continually increases, data intensive quantitative analyses will increase in use. Approaches such as co-occurrance, network modeling, and expert networks described above will become more commonplace in research assessment.

For those fields of technology in which patents are an important mode of communication, patent citation analysis offers insight into the conversion of science to technology. Many of the reported patent citation analysis studies tend to focus on technical intelligence for corporate applications (*Narin*, 1994).

## III. Research requirements for RIA summary

More retrospective studies are required using modern technologies such as information processing and computerized citation databases. The tracing of the indirect impacts of research should be emphasized. Network approaches are valuable in this regard. More rigorous peer review experiments should be performed, to understand better the issues of cost, validity, reliability, quality, and feedback. The text describes the main parameters to be examined in these studies. For bibliometrics, studies are required to address the normative comparisons across different disciplines, as well as to examine optimal ways to combine multiple indicators into few figures of merit.

Central to the assessment of research is the capability to handle all phases of the information creation, flow, and integration cycle. The explosion of available information in the last decade requires the utilization of large databases to handle this information in support of RIA.

In particular, sophisticated data collection, analysis, and interpretation schemes can track the dissemination of information flowing from research to other applications. A credible research product tracking scheme can help identify the indirect impacts of research more precisely, and can improve correlations between research evaluation predictions (such as peer review and bibliometrics) and downstream impacts.

Central to credible work in predicting and tracking the diffusion of information from research is a database of research products at various evolutionary stages which can feed the predictive models. This database of research products could be linked in part with databases of sponsored research and technology. Since the research product evolutionary pathways transcend the research originating organization, and can intersect all societal sectors, the cooperation of many public and private organizations would be required to develop a database of research products in their evolutionary stages. Development and construction of such a database should start now.

Comprehensive databases describing sponsored research and development programs in many funding agencies and organizations, with sophisticated software to provide rapid access to the database contents, can help improve the selection, management, and evaluation of research programs. Research gaps can be identified, duplication of programs can be minimized, complementary and joint programs can be established, substantial leveraging of other agency programs can be implemented, and technology planning can be improved with better awareness of maturing research programs.

To fully understand a research program, especially in the assessment of that program, evaluators must be cognizant of the large body of research being conducted throughout the world. In addition, to fully understand the impacts of research on different technologies, evaluators must be cognizant of the large body of existing and developmental technology throughout the world, and the existing and potential shortcomings in those technologies.

With the advent of high speed and high storage capacity computers, and advances in database software packages, the capability exists now to make large amounts of information available to researchers and evaluators. In particular, the capability exists to provide information about funded research and technology development programs being conducted throughout the world, as well as information about existing technologies.

Tailored databases which contain information about the structural relationships among projects and programs can help identify critical paths for development in R&D programs. This is important in allocating resources among programs in mission-oriented agencies and other organizations.

Sophisticated algorithms for manipulating and interpreting large technical textual databases would allow pervasive themes of the databases to be identified, as well as the relationships among the themes and sub-themes. Low frequency anomolous relationships which could be important are identified easily with these techniques. The algorithms would also allow identification of the translations between research areas and technology areas in the databases, and would provide guidelines and roadmaps for increasing the efficiency of searching unfamiliar databases.

These algorithms, and subsequent analyses, have the potential of identifying emerging research and development areas contained within the databases but not readily discernable. The software can also help in taxonomy construction, with the taxonomy elements obtained 'bottom-up' from the database language, rather than top down using an authoritative directed approach. Many different types of taxonomies could be constructed from the full text database, and relationships among the different elements of the different taxonomies could be obtained. Finally, by looking at the changes in the structure of research fields over time, the impact of sponsoring organization intervention can be ascertained.

## IV. Bibliography

1. BROWN, G. E., "Report of the Task Force on the Health of Research," Chairman's Report to the Committee on Science, Space, and Technology, U.S. House of Representatives, No. 56-819, U.S. Government Printing Office, Washington, 1992.
2. NAS, "The Government Role in Civilian Technology: Building a New Alliance", Committee on Science, Engineering, and Public Policy, National Academy of Sciences, National Academy Press, 1992.
3. Carnegie, "Enabling the Future: Linking Science and Technology to Societal Goals," Carnegie Commission on Science, Technology, and Government, Carnegie Commission, New York, NY, 1992.
4. OTA, "Federally Funded Research: Decisions for a Decade", U.S.Congress, Office of Technology Assessment, OTA-SET-490 (Wash., DC: U. S. GPO, May 1991).
5. OTA, "The Defense Technology Base: Introduction and Overview", U.S. Congress, Office of Technology Assessment, (OTA-ISC-374, March 1988) and "Holding the Edge: Maintaining the Defense Technology Base" (OTA-ISC-420) Wash., DC: U. S. GPO, April 1989.

6. NARIN, F., "The Impact of Different Modes of Research Funding", in: EVERED, D., HARNETT, S. (Eds.), *The Evaluation of Scientific Research*, John Wiley and Sons, Chichester, UK, 1989.
7. ROBB, W. L., "Evaluating Industrial R&D", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
8. NELSON, K. S., TOMSYCK, J. P., SORENSEN, D. P., "Industrial R&D Program Evaluation Techniques", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
9. SALASIN, J. et al, "The Evaluation of Federal Research Programs", MITRE Technical Report MTR-80W123, June 1980.
10. LOGSDON, J.M., RUBIN, C.B., "An Overview of Federal Research Evaluation Activities", Report, The George Washington University, Wash., D. C., April 1985. See also J. M. LOGSDON, C. B. RUBIN, *Federal Research Evaluation Activities*, Cambridge, MA, Abt Associates, 1985.
11. CHUBIN, D. E., HACKETT, E. J., *Peerless Science: Peer Review and U. S. Science Policy*, State University of New York Press, Albany, NY, 1990.
12. CHUBIN, D. E., "Grants Peer Review in Theory and Practice", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
13. KOSTOFF, R. N., "Evaluating Federal R&D in the U. S.," in: *Assessing R&D Impacts: Method and Practice*, BOZEMAN, B., MELKERS, J. (Eds), Kluwer Academic Publishers, Norwell, MA, 1993.
14. KOSTOFF, R. N., "Quantitative/Qualitative Federal Research Impact Evaluation Practices", *Technological Forecasting and Social Change*, 45:2, February 1994.
15. KOSTOFF, R. N., "Research Impact Assessment: Federal Peer Review Practices", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
16. BARKER, K., "The 'British Model' – Evaluation by Professionals", in: LAREDO, P., MUSTAR, P. (Eds), *EC Handbook on Evaluation*, 1992.
17. CICCHETTI, D. V., "The Reliability of Peer Review for Manuscript and Grant Submissions: A Cross-Disciplinary Investigation," *Behavioral and Brain Sciences*, 14:1, 1991.
18. COLE, S., RUBIN, L., COLE, J., "Peer Review in the National Science Foundation: phase one of a study,", National Research Council, 1978, NTIS Acc. No. PB83-192161.
19. COLE, J., COLE, S., "Peer Review in the National Science Foundation: phase two of a study," National Research Council, 1981, NTIS Acc. No. PB82-182130.
20. COLE, S., COLE, J., SIMON, G., "Chance and Consensus in Peer Review," *Science*, Vol. 214, November 1981.
21. COZZENS, S. E., "Expert Review in Evaluating Programs", *Science and Public Policy*, 14:2, April 1987.
22. DOD, "The Department of Defense Report on the Merit Review Process for Competitive Selection of University Research Projects and an Analysis of the Potential for Expanding the Geographic Distribution of Research," April 1987, DTIC Acc. No. 88419044.
23. DOE, "An Assessment of the Basic Energy Sciences Program", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0123, March 1982.
24. DOE, "Procedures for Peer Review Assessments", Office of Energy Research, Office of Program Analysis, Report No. DOE/ST-0007P, Revised January 1993.
25. FRAZIER, S. P., "University Funding: Information on the Role of Peer Review at NSF and NIH", U.S. General Accounting Office Report No. GAO/RCED-87-87FS, March 1987.
26. KOSTOFF, R. N., "Evaluation of Proposed and Existing Accelerated Research Programs by the Office of Naval Research", *IEEE Trans. of Engineering Management*, 35:4, Nov. 1988.
27. ORMALA, E., "Nordic Experiences of the Evaluation of Technical Research and Development", *Research Policy*, 18, 1989.
28. OTA, "Research Funding as an Investment: Can We Measure the Returns", U. S. Congress, Office of Technology Assessment, OTA-TM-SET-36 (Wash., DC: U. S. GPO, April 1986).
29. NICHOLSON, R. S., "Improving Research Through Peer Review," National Research Council, 1987, NTIS Acc. No. PB88-163571.
30. DOE, "An Evaluation of Alternate Magnetic Fusion Concepts 1977", DOE/ET – 0047, May 1978.

31. NIST, "Annual Report, 1990," Visiting Committee on Advanced Technology, January 1991.
32. ORMALA, E., "Impact Assessment: European Experience of Qualitative Methods and Practices", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
33. ROY, R., "Funding Science: The Real Defects of Peer Review and an Alternative to It", *Science, Technology, and Human Values*, 10:3, 1985.
34. KING, J., "A Review of Bibliometric and Other Science Indicators and Their Role in Research Evaluation", *Journal of Information Science*, 13, 1987.
35. KRUYTBOSCH, C., "The Role and Effectiveness of Peer Review", in: EVERED, D., HARNETT, S. (Eds.), *The Evaluation of Scientific Research*, John Wiley and Sons, Chichester, UK, 1989.
36. BORNSTEIN, R. F., "The Predictive Validity of Peer Review: A Neglected Issue," *Behavioral and Brain Sciences*, 14:1, 1991.
37. BORNSTEIN, R. F., "Manuscript Review in Psychology: Psychometrics, Demand Characteristics, and an Alternative Model," *Journal of Mind and Behaviour*, 12, 1991.
38. NARIN, F., OLIVASTRO, D., STEVENS, K. A., "Bibliometrics -Theory, Practice, and Problems", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
39. MANSFIELD, E., "Academic Research and Industrial Innovation," *Research Policy*, Vol. 20, 1991.
40. KOSTOFF, R. N., "Semi-Quantitative Methods for Research Impact Assessment", *Technological Forecasting and Social Change*, 44:3, November, 1993.
41. KINGSLEY, G., "The Use of Case Studies in R&D Impact Evaluations", in: *Assessing R&D Impacts: Method and Practice*, BOZEMAN, B., MELKERS, J. (Eds.), Kluwer Academic Publishers, Norwell, MA, 1993.
42. DOD, *Project Hindsight*, Office of the Director of Defense Research and Engineering, Wash., D. C., DTIC No. AD495905, Oct. 1969.
43. IITRI, "Technology in Retrospect and Critical Events in Science", Illinois Institute of Technology Research Institute Report, December, 1968.
44. Battelle, "Interactions of Science and Technology in the Innovative Process: Some Case Studies", Final Report, Prepared for the National Science Foundation, Contract NSF-C 667, Battelle Columbus Laboratories, March 19, 1973.
45. IDA, "DARPA Technical Accomplishments", Volume I, IDA Paper P-2192, February 1990; Volume II, IDA Paper P-2429, April 1991; Volume III, IDA Paper P-2538, July 1991, Institute for Defense Analysis.
46. DOE, "Health and Environmental Research: Summary of Accomplishments", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0194, May 1983
47. DOE, "Health and Environmental Research: Summary of Accomplishments", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0275, August 1986.
48. KOSTOFF, R. N., "Research Impact Quantification," *R&D Management*, 24:3, July 1994.
49. Australia, "Research Performance Indicators Survey", National Board of Employment, Education and Training, Commissioned Report No. 21, Australian Government Publishing Service, Canberra, Australia, January 1993.
50. BRAUN, T., GLÄNZEL, W., SCHUBERT, A., "An Alternative Quantitative Approach to the Assessment of National Performance in Basic Research", in: EVERED, D., HARNETT, S. (Eds.), *The Evaluation of Scientific Research*, John Wiley and Sons, Chichester, UK, 1989.
51. BRAUN, T., et al, "Publication Productivity: from Frequency Distributions to Scientometric indicators," *Journal of Information Science*, 16, 1990.
52. BRAUN, T., et al, "Scientometric Indicators Datafiles," *Scientometrics*, 28:2, 1993.
53. SCHUBERT, A., BRAUN, T., "Relative Indicators and Relational Charts for Comparative Assessment of Publication Output and Citation Impact," *Scientometrics*, 9:5-6, 1986.
54. BRAUN, T., SCHUBERT, A., "Scientometric versus Socio-Economic Indicators: Scatter Plots for 51 Countries: 1978-1980," *Scientometrics*, 13:1-2, 1987.

55. BRAUN, T., SCHUBERT, A., "The Landscape of National Performances in the Sciences, 1981-1985," *Scientometrics*, 20:1, 1991.
56. SCHUBERT, A., BRAUN, T., "Three Scientometric Etudes on Developing Countries as a Tribute to Michael Moravcsik," *Scientometrics*, 23:1, 1992.
57. OBERSKI, J. E. J., "Some Statistical Aspects of Co-citation Cluster Analysis and a Judgement by Physicists," in: VAN RAAN, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland, 1988.
58. WHITE, H.D., MCCAIN, K.W., "Bibliometrics," in: WILLIAMS, M.E. (Ed.), *Annual Review of Information Science and Technology*, 24, 1989.
59. NARIN, F., "Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity" (monograph), NSF C-637, National Science Foundation, Contract NSF C-627, NTIS Accession No. PB252339/AS, March 31, 1976.
60. HICKS, D., MARTIN, B., IRVINE, J., "Bibliometric Techniques for Monitoring Performance in Technologically Oriented Research: The Case of Integrated Optics", *R&D Management*, Vol. 16, No. 3, 1986.
61. NSF, "Science and Engineering Indicators – 1989", National Science Board Report NSB 89-1, GPO, Wash., D.C., 1989.
62. MARTIN, B. R. et al, "Recent Trends in the Output and Impact of British Science", *Science and Public Policy*, 17:1, Feb., 1990.
63. FRAME, J. D., "Quantitative Indicators for Evaluation of Basic Research Programs/ Projects", *IEEE Transactions on Engineering Management*, Vol. EM-30, No. 3, August 1983.
64. MCALLISTER, P.R., NARIN, F., CORRIGAN, J.G., "Programmatic Evaluation and Comparison Based on Standardized Citation Scores", *IEEE Transactions on Engineering Management*, Vol. EM-30, No. 4, November 1983.
65. MULLINS, N., "Evaluating Research Programs: Measurement and Data Sources", *Science and Public Policy*, Vol. 14, No. 2, April 1987.
66. MULLINS, N., SNIZEK, W., OEHLER, K., "The Structural Analysis of a Scientific Paper," in: VAN RAAN, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland, 1988
67. MOED, H.F., VAN RAAN, A.F.J., "Indicators of Research Performance: Applications in University Research Policy," in: VAN RAAN, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland, 1988
68. IRVINE, J., "Evaluation of Scientific Institutions: Lessons from a Bibliometric Study of UK Technical Universities", in: EVERED, D., HARNETT, S. (Eds.), *The Evaluation of Scientific Research*, John Wiley and Sons, Chichester, UK, 1989.
69. VAN RAAN, A.F.J., "Evaluation of Research Groups", in: EVERED, D., HARNETT, S. (Eds.), *The Evaluation of Scientific Research*, John Wiley and Sons, UK, 1989.
70. LUUKKONEN, T., "Bibliometrics and Evaluation of Research Performance", *Annals of Medicine*, Vol. 22, No. 3, 1990.
71. LUUKKONEN, T., STAHLE, B., "Quality Evaluations in the Management of Basic and Applied Research", *Research Policy*, 19, 1990.
72. LUUKKONEN, T., PERSSON, O., SIVERTSEN, G., "Understanding Patterns of International Scientific Collaboration," *Science, Technology, and Human Values*, Vol. 17, No. 1, January 1992.
73. NARIN, F., "Bibliometric Techniques in the Evaluation of Research Programs", *Science and Public Policy*, 14:2, April 1987.
74. CARPENTER, M. P., NARIN, F., "Validation Study: Patent Citations as Indicators of Science and Foreign Dependence", *World Patent Information*, Vol. 5, No. 3, 1983.
75. NARIN, F., CARPENTER, M. P., WOOLF, P., "Technological Performance Assessments Based on Patents and Patent Citations", *IEEE Transactions on Engineering Management*, EM-31, 4, Nov. 1984.
76. WALLMARK, J.T., SEDIG, K.G., "Quality of Research Measured by Citation Method and by Peer Review – A Comparison", *IEEE Transactions on Engineering Management*, Vol. EM-33, No. 4, November 1986.

77. COLLINS, P., WYATT, S., "Citations in Patents to the Basic Research Literature", *Research Policy*, 17, 1988.
78. NARIN, F., OLIVASTRO, D., "Technology Indicators Based on Patents and Patent Citations", in: VAN RAAN, A.F.J. (Ed.), *Handbook of Quantitative Studies in Science and Technology*, Elsevier Science Publishers, Amsterdam, 1988.
79. VAN VIANEN, B. G., MOED, H. F., VAN RAAN, A. F. J., "An Exploration of the Science Base of Recent Technology," *Science Policy*, Vol. 19, 1990.
80. NARIN, F., OLIVASTRO, D., "Status Report-Linkage between Technology and Science," *Research Policy*, 21:3, June, 1992.
81. CARPENTER, M.P., COOPER, M., NARIN, F., "Linkage Between Basic Research Literature and Patents", *Research Management*, 13:2, March 1980.
82. NARIN, F., NOMA, E., PERRY, R., "Patents as Indicators of Corporate Technological Strength", *Research Policy*, Vol. 16, 1987.
83. NARIN, F., "Technological Evaluation of Industrial Firms by Means of Patent Investigation", Presented at VPP Professional Meeting, Nürnberg, Germany, November 13, 1992.
84. MILLER, R., "The Influence of Primary Task on R&D Laboratory Evaluation: A Comparative Bibliometric Analysis", *R&D Management*, 22:1, 1992.
85. SCHUBERT, A., BRAUN, T., "Reference Standards for Citation Based Assessments", *Scientometrics*, 26:1, 1993.
86. KOSTOFF, R. N., "Research Impact Assessment," *Proceedings: Third International Conference on Management of Technology*, Miami, FL, February 17-21, 1992. Larger text available from author.
87. KOSTOFF, R. N., "Co-Word Analysis," in: *Assessing R&D Impacts: Method and Practice*, BOZEMAN, B., MELKERS, J. (Eds.), Kluwer Academic Publishers, Norwell, MA, 1993.
88. KOSTOFF, R.N., "Database Tomography: Origins and Applications," *Competitive Intelligence Review*, Special Issue on Technology, 5:1, Spring 1994.
89. TIJSSEN, R., VAN RAAN, A., "Mapping Changes in Science and Technology", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
90. GEORGHIOU, L., GIUSTI, W.L., CAMERON, H.M., GIBBONS, M., "The Use of Co-nomination Analysis in the Evaluation of Collaborative Research," in: VAN RAAN, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland, 1988.
91. ENGELSMAN, E. C., VAN RAAN, A. F. J., "Mapping of Technology: A First Exploration of Knowledge Diffusion amongst Fields of Technology," Research Report to the Ministry of Economic Affairs, CWTS-91-02, Centre for Science and Technology Studies, Leiden, March 1991.
92. AVERCH, H., "Economic Approaches to the Evaluation of Research", in: KOSTOFF, R. N. (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994.
93. LINK, A., "Methods for Evaluating the Return on R&D Investments", in: *Assessing R&D Impacts: Method and Practice*, BOZEMAN, B., MELKERS, J. (Eds.), Kluwer Academic Publishers, Norwell, MA, 1993.
94. ROESSNER, J. D., "Use of Quantitative Methods to Support Research Decisions in Business and Government", in: *Assessing R&D Impacts: Method and Practice*, BOZEMAN, B., MELKERS, J. (Eds.), Kluwer Academic Publishers, Norwell, MA, 1993.
95. KOSTOFF, R. N., "A Cost/ Benefit Analysis of Commercial Fusion-Fission Hybrid Reactor Development", *Journal of Fusion Energy*, 3:2, 1983.
96. MANSFIELD, E., "Basic Research and Productivity Increase in Manufacturing," *The American Economic Review*, Vol. 70, No. 5, December 1980.
97. TERLECKYJ, N., *State of Science and Research: Some New Indicators*, Westview Press, Boulder, CO, 1977.
98. TERLECKYJ, N., "Measuring Economic Effects of Federal R&D Expenditures: Recent History with Special Emphasis on Federal R&D Performed in Industry", Presented at NAS Workshop on 'The Federal Role in Research and Development', November 1985.

99. GRILICHES, Z., "Issues in Assessing the Contribution of Research and Development to Productivity Growth", *The Bell Journal of Economics*, Vol. 10, Spring 1979.
100. GRILICHES, Z., "Productivity, R&D, and the Data Constraint", *The American Economic Review*, 84:1, March 1994.
101. AVERCH, H., "Measuring the Cost-Efficiency of Basic Research Investment: Input-Output Approaches", *Journal of Policy Analysis and Management*, Vol. 6, No. 3, 1987.
102. AVERCH, H., "Exploring the Cost-Efficiency of Basic Research Funding in Chemistry", *Research Policy*, Vol. 18, 1989.
103. ODEYALE, C.O., 1993. *Knowledge-Based Systems: Knowledge Representation and Inference Strategies for Effective and Unbiased Military Biomedical and R&D Management.* Ph.D. Thesis, Walden Univ.
104. ODEYALE, C. O., KOSTOFF, R. N., "R&D Management Expert Networks: I. Knowledge Representation and Inference Strategies", *HEURISTICS, the Journal of Knowledge Engineering and Technology*, 7:1, 1994.
105. ODEYALE, C. O., KOSTOFF, R. N., "R&D Management Expert Networks: II. Prototype Construction and Validation", *HEURISTICS, the Journal of Knowledge Engineering and Technology*, 7:1, 1994.

## V. Suggestions for further reading

VAN RAAN, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North Holland, 1988. This compendium, assembled by one of the leading research evaluators in the world, focuses on quantitative measures of S&T output.

KOSTOFF, R. N., (Ed.), *Evaluation Review*, Special Issue on Research Impact Assessment, 18:1, February 1994. This special issue addresses the quantitative, retrospective, and qualitative approaches for assessing research, and examines research evaluation from government, industrial, and European viewpoints.

BOZEMAN, B., MELKERS, J. (Eds), *Assessing R&D Impacts: Method and Practice*, Kluwer Academic Publishers, Norwell, MA, 1993. This comprehensive book examines the many facets of research evaluation.

KOSTOFF, R. N., *Handbook of Research Impact Assessment*, 1995. Available from author. In its latest incarnation (~425 pages, ~2300 item bibliography), this Handbook examines the different research assessment approaches, and recommends evaluation procedures for Federal agencies.