NISTEP STUDY MATERIAL

No. 86

Quantitative Methods of Research Evaluation Used by the U.S. Federal Government

Diana Hicks, Peter Kroll, Francis Narin,

Patrick Thomas, Rosalie Ruegg,

Hiroyuki Tomizawa, Yoshiko Saitoh, and Shinichi Kobayashi

Second Theory-Oriented Research Group,
National Institute of Science and Technology Policy

May 2002

This research was conducted by National Institute of Science and Technology Policy (NISTEP) in cooperation with CHI Research, Inc. Among writers, Diana Hicks, Peter Kroll, Francis Narin and Patrick Thomas belong to CHI Research, Inc. and Rosalie Ruegg belongs to Technology Impact Assessment. The compiling work of this reference material was, however, conducted as a task of NISTEP. During the investigation, we received cooperation from Japan Science and Technology Corporation.

Quantitative Methods of Research Evaluation
Used by the U.S. Federal Government

May 2002

Diana Hicks†, Peter Kroll†, Francis Narin†, Patrick Thomas†,
Rosalie Ruegg‡, Hiroyuki Tomizawa, Yoshiko Saitoh,
and Sinichi Kobayashi
(† CHI Research, Inc., ‡ Technology Impact Assessment)

Second Theory-Oriented Research Group,
National Institute of Science and Technology Policy (NISTEP)
Ministry of Education, Culture, Sports, Science and Technology (MEXT)
JAPAN

Preface

The purpose of this report is to describe various quantitative methods used in research evaluations in the United States, for use as a reference in quantitative evaluations to be conducted in Japan. Reviewed in this report are quantitative methods adopted for R&D programs/projects by the U.S. federal government, and focusing, in particular, on economic and bibliometrical analyses used for evaluating the results and the outcomes of programs and projects.

The mainstream of analytical methods in the United States has traditionally been peer reviews, where evaluations are conducted by experts who can understand the contents of a study well. Extensive application of quantitative evaluation methods is, therefore, a relatively recent trend. In the background to such a shift lies a situation that allowed the introduction of an administrative management system that attaches great importance to records and results, and a tendency that values the accountability of the government to the nation. The research evaluation methods reviewed in this report have gradually developed under such circumstances through constant calls for higher accuracy, as well as severe criticism.

In the meantime, the research evaluation system has also been improved in Japan based on the needs for more efficient administration and to gain the proper understanding of the people, as financial support for research and development activities have been reinforced. Amid a situation where the entire Japanese administration has embarked on policy evaluations, science and technology-related administration has also started evaluating R&D programs and policies themselves, rather than merely evaluating the objects of the policy. Although quantitative methods are not almighty and their applications are limited, they are now an indispensable tool in conducting evaluations in the United States. On the contrary, economic analyses as well as bibliometrical analyses have almost no accumulated experiences in Japan, and cases of their practical application are still quite rare.

It is often said that there is no general evaluation method that is applicable to all cases and that an applicable method should be engineered carefully for each evaluation case. Therefore, it is necessary to seek the best application for each evaluation method, based on a good understanding of the characteristics of individual methods and the circumstances where each analysis was used in the United States, rather than just introducing analytical methods from the United States to Japan.

In compiling this report, we aimed at providing as much diversified information as possible including the context of the use of each quantitative evaluation method developed in the United States, in addition to explaining technical aspects of the individual methods. We sincerely hope that this report can provide useful suggestions for research evaluations in Japan.

Notes for the English Version

This report was compiled by NISTEP with the cooperation of CHI Research, Inc. For this work, besides the staff members of CHI Research, Inc. that has experiences in research and development activities in the United States, Ms. Rosalie Ruegg who had been involved in research evaluation at ATP also took part. Dr. Diana Hicks of CHI Research, Inc., who has a good understanding of the needs on the Japanese side, based on her works at NISTEP in the past as a visiting researcher, played successfully a bridging role between NISTEP and CHI Research, Inc.

This English version of the report was compiled on the U. S. side based on the draft report made through discussion between NISTEP and CHI Research, Inc. While the Japanese version includes some explanations and implications for the Japanese readers, such supplemental items are omitted in this English version.

Table of Contents

"Quantitative Methods of Research Evaluation Used by the U.S. Federal Government"

Executive Summary

This report describes quantitative techniques for research evaluation as used in the United States. At the heart of the report are descriptions of evaluations using economic and bibliometric methodology. The scene is set in the first chapter which provides an historical overview of research evaluation at the Federal level with commentary on recent trends and on the place of quantitative evaluation within the system.

Economic methods are described in Chapter 2 by Rosalie Ruegg, former Director of the Economic Assessment Office of the Advanced Technology Program (ATP), National Institute of Standards (NIST). In the late 1990s, ATP's ongoing evaluation program commissioned a series of ever more methodologically sophisticated evaluations from some of the nation's leading economists concerned with science, technology and innovation. This unique body of work is quite accessible in the public literature and illustrates the uses of economic methods in R&D program and project evaluation.

Bibliometric methods are described in Chapter 3, written by Diana Hicks and Francis Narin, and Chapter 4 written by Diana Hicks, Peter Kroll and Patrick Thomas of CHI Research. For two decades CHI Research has produced bibliometric evaluations for the Federal government. CHI works with three types of citation data: paper-paper, patent-patent and patent-paper. All three methods are described. Chapter 3 presents the development and validation of the methods, their strengths and weaknesses and the elements of bibliometric method necessary to produce evaluations using databases of papers and patents. Chapter 4 presents summaries of CHI's bibliometric work for the Federal government as a means of illustrating both the variety of questions that can be addressed and how the techniques introduced in Chapter 3 are implemented in practice. Chapter 4 also presents summaries of recent notable evaluations by others, most of which incorporate bibliometric methods.

Chapter I. Overview of the Evaluation of Publicly Funded R&D in the United States

by Diana Hicks

Several themes are apparent in the American literature on post-project or program research evaluation at the Federal level. Peer review methods are far more prevalent than quantitative methods in Washington and have a strong advocate/contractor in the National Academies. Quantitative evaluation is said to be not much used here, and this judgment appears to be formed in comparison with the European situation where evaluation has been commonplace for several decades. The lack of use is lamented, reasons are adduced, and the belief that peer review and quantitative methods should be used together is expressed. There is also a high level of fragmentation in the area, agencies often do not know about all the studies they have commissioned and reviews seem to focus on works in which the author was directly involved. Finally, evaluations and evaluation methods are often contested. Evaluation results feed policy arguments that swirl around resource allocation, and because the U.S. policy-making system is reasonably open, there are many parties to these arguments. Furthermore, many analyses are performed by outside contractors - consulting companies or academics - who compete for contracts, and who tend to advocate their own methods at the expense of others.
This chapter will discuss these themes, providing pointers to the literature along the way. Given the nature of the situation, this chapter and the report as a whole are shaped by the author's position which should be made clear. CHI Research, Inc. has performed bibliometric evaluations for various agencies of the U.S. Federal government for two decades. Therefore, we have insights that go beyond the published literature, but nevertheless, the fragmented situation limits the comprehensiveness of the overview that we can present. This review focuses on our work, in line with most reviews in the U.S. literature. Because we perform bibliometric evaluations, we can and do put forward a strong case for bibliometrics and explain bibliometric methodology at a fairly sophisticated level. This is not characteristic of most US literature on research evaluation.

The fundamental nature of evaluations

For the purposes of this discussion, we will classify evaluations as internal or external, based on who conducts the evaluation. By internal evaluation we mean studies performed entirely by the scientific community, whether the experts are drawn just from within an agency or whether outside scientific expertise is used. Internal evaluation most often relies upon qualitative methods like peer review or anecdote – that is compiling books describing program accomplishments. External evaluation draws on some sort of social scientific expertise, such as economics, bibliometrics, or case study methodology and quantitative methods are much more likely to be used. External expertise is indeed usually found outside the agency, though expertise external to the scientific community need not always be external to an agency. For example, ATP uses in-house economists for evaluation, and consultants can be hired to compile books of accomplishments. Therefore, the classification is not perfectly clean, yet it holds in most cases.

In the United States, internal methods are used most frequently - particularly peer review. In the mid-1980s, Logsdon and Rubin (1988) interviewed 44 people responsible for research management and research evaluation in 10 Federal agencies. They reported widespread use of peer review, with almost every agency they contacted using the technique. Nothing suggests that things have changed since the late 1980s, though nobody seems to have repeated their research. Logsdon and Rubin noted that this situation is in line with the recommendations the National Academies made in a 1982 report examining evaluation methods. Not infrequently, the National Academies were contracted to perform the peer reviews. Non-quantitative case studies and anecdotal evaluations, the reporting of "nuggets" illustrating recent agency accomplishments are also widely used by agencies according to Kostoff (1993).

As for external methods, one form of semi-quantitative case study has a distinguished history in the U.S. This is the tracing methodology in which innovations are chosen and their antecedent research is identified. Project Hindsight from the Department of Defense (DOD, 1969) was the first such study. It looked at the research background of weapons systems and found that ideas leading to enhanced weapons systems occurred when research scientists or engineers were intimately aware of problems of the applications engineer. The National Science Foundation followed up with the TRACES study (IITRI, 1968) which traced back several innovations, but went back further in time than the Hindsight study, and found significant contributions from basic research. Over the years, more studies followed including "TRACES II" (Battelle, 1973), and a tracing of the underpinnings of advances in cancer care, and most recently a tracing of NSF contributions to engineering innovations which is summarized later in Chapter 4 (Roessner et al., 1997, 1998b).

The strengths and weaknesses of such studies are well understood. Georghiou and Roessner (2000) report that the method provides rich information about the relative contribution of basic vs. applied research, institutional contexts and sources of research support. But much judgment about the significance of research events is required to construct a tracing. Because technological antecedents are so complex, tracings focus on precursors judged to be "significant", thus weaker links and dead ends are dropped and the picture produced ignores indirect effects. The method is also expensive and so studies are limited to a handful of innovations not chosen at random, therefore the results cannot be generalized. The method is not suitable for evaluating the outcome of a program or project (because it traces backwards, not forwards). It is better at providing agencies with material to use in advocacy because it demonstrates the importance of basic science to the advance of technology.

The quantitative methods used in external evaluation are divided into two broad types - economic and bibliometric analysis. These methods are the topic of this report which will explain what has been achieved using these techniques in America and will explore the strengths and weakness of the techniques. Chapter 2 examines economic methods used in evaluations of the Advanced Technology Program (ATP) of the National Institute of Standards and Technology. ATP has developed these methods to their highest level through its commissioning of evaluations. Bibliometrics is the subject of Chapters 3 and 4. Chapter 3 explains the elements needed to use bibliometrics for evaluation while Chapter 4 comprises summaries of evaluations. Most of these are bibliometric evaluations conducted by CHI for the Federal government over the past two decades. There are however four summaries illustrating non-CHI, external evaluations: the tracing done for NSF by Roessner et al. mentioned above; a large case study evaluation for the Department of Energy by Bozeman et al. (research value mapping project); interactive data mining/bibliometrics for the Navy by Kostoff; and international benchmarking as a methodological experiment undertaken by COSEPUP of the National Academies.[1] These are examples of more qualitative evaluations, though all contain a bibliometric component.

[1] The COSEPUP study would be classified as “internal”, the rest as external.

The infrequency with which non-peer review methods – i.e. external methods - are used in the United States is a theme found in all reviews of research evaluation. Kostoff (1993) notes that there are many large bibliographies of papers reporting methods developed to evaluate research conduct, impact and benefits. A relatively small fraction of the methods are actually implemented by Federal agencies. Of those implemented, only a small fraction of the results are published, and an even smaller fraction are accepted by the decision-makers. Kostoff orders methods by frequency of use in the following way: peer review, non-quantitative case study and anecdotal, and quantitative. This report dispenses with proposals and ideals, discussing instead methods that have been used. This cuts down considerably on the amount of material that needs to be covered.

The focus on the least frequently used methods – quantitative – further cuts down the material. Yet, there remains considerable ground to cover because both economic and bibliometric studies have been commissioned by U.S. Federal research funding agencies for several decades. The Office of Technology Assessment reviewed research evaluation in 1985 and reported several experiments at NASA using economic methods of analysis. In the late 1980s, Logsdon and Rubin report that the National Bureau of Standards of the Department of Commerce had conducted studies of research outcomes using microeconomic techniques. Curiously, today the NBS's successor, the National Institute of Standards and Technology has developed economic methods for research evaluation to perhaps their highest level – see Chapter 2.

Bibliometric methods are also infrequently used in the United States. Bibliometric methods were extensively covered in the 1985 OTA report. The OTA, Logsdon and Rubin and others at the time reported at length on CHI's work with the National Institute of Health (NIH). These studies are summarized in this report. Logsdon and Rubin also found extensive use of bibliometrics, half of the agencies examined used bibliometrics in some form (9 of 18) although more used peer review (17 of 18) (Logsdon & Rubin, 1988, Table 2). In a second report in 1991, the OTA found little advance in methods and focused more on lack of use and how studies should be structured. Both studies concluded that quantitative methods were not much used in the U.S. This judgment was no doubt made in comparison with Europe, where the techniques are pervasive. OTA 1985 included pages of discussion of British bibliometrics and a bibliography compiled for OTA 1991 contains a large European component (Averch, 1993a).

In recent years, research evaluation at the Federal level has increased. Melkers and Roessner (1997) point to a “rebirth” of evaluative activities both at the Federal and state levels. CHI's records support this assessment, suggesting that since the OTA reports, commissioning of studies has increased considerably. Figure 1 reports the number of studies conducted by CHI Research for agencies of the Federal government in three time periods, 1977-84, 1985-92 and 1993-00. The timing of the OTA reviews is noted. There is further evidence for an increase in non-peer review evaluation activity. The non-CHI studies summarized in Chapter 4 all were conducted in the late 1990s. In addition a record of the work of Abt Associates, a consultancy that has undertaken historical case studies for research evaluation, is available on their website which lists studies conducted for the National Science Foundation. Before 1991, there were several reviews of methods of evaluation, and some quantitative analyses of university publishing rates and other science policy relevant work. In the late 1990s there were five studies, all of whose titles began: "An evaluation of . . ." Several hypotheses are available to explain the apparent increase in non-peer review activity.

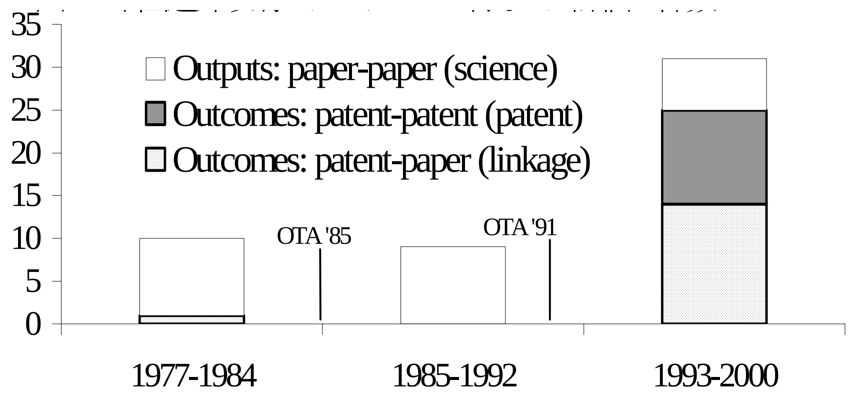


Figure I-1. Number of evaluations conducted by CHI for the US federal government

The first, and best supported, hypothesis is that the political climate changed (Melkers and Roessner, 1997). In the beginning, this was more noticeable abroad. But in 1991 the OTA concluded that demand for research evaluation had increased even in the United States in response to budgetary constraints, greater accountability to sponsors, and the desire for increased rationality in decision making. They stated that many agencies in the U.S. had formed evaluation units (OTA, 1991, p. 252-253). These trends strengthened as the decade wore on. In 1995, Cozzens remarked on the increasing pressure on the discretionary portion of the Federal budget (where research funding comes from) and on congressional action on their mandate to keep government spending under control. The theme of greater accountability noted by OTA was also identified by Cozzens. She noted: "A newly serious attitude has emerged about examining the benefits to the American public of various programs." In 1993 Senator Barbara Milulski sent shock waves through the research community by asking the National Science Foundation to link the majority of its programs to national goals, to set milestones for moving toward the goals, and to report back on its progress. Cozzens sees this as part of renegotiating the "science/society contract" in the post-Cold war era (Cozzens, 1995b, p. 352). These more stringent conditions are precisely the sort of thing that led to increased non-peer review evaluation in Europe a decade earlier.

Second, there was a growing emphasis on evaluating the outcomes from research, as opposed to outputs. Research outputs are the traditional products of research such as new knowledge and scientific excellence as manifested in publications and scientific reputations. The term "outcome" is used to mean the effect of research on society or the economic benefits, new technologies, environmental improvements etc. that the country gains from its investments in research. Evaluation of outputs has been the strength of peer review; the movement to outcomes sidelines peers in favor of users. Similarly, traditional paper-paper bibliometrics is seen as somewhat irrelevant to the evaluation of research outcomes (Georghiou and Roessner (2000) quoting Cozzens et al. 1994.)

These days, research for the sake of research is of considerably less interest than is research more directly connected with technology or other application. In part this represents an evolution in the nature of research so that today much more research is much more closely linked to technology, in the life sciences or information technology for example. Narin has used evidence based on patterns of patents citing papers to argue that science and technology are becoming intermingled (Narin & Noma, 1985). Stokes argued that science policy needs to get beyond the basic/applied distinction because so much work today is both basic and relevant to application, as was Pasteur's work (Stokes, 1997). CHI's examination of papers cited in patents supports Stokes argument in that the work cited is in basic journals, and tends to be highly cited in scientific papers as well. *Nature* and *Science* are two of the journals most intensively cited in patents (Hicks et al., 2000). The increase in top quality research with close links to outcomes makes it easier to demonstrate those links. This has fairly direct consequences for evaluation in that the increasing rate of citation from U.S. patents to papers has made more visible the connections between research and technology. Patent-paper bibliometrics has thus become available as an outcome-related evaluation tool, and patent-paper bibliometrics account for a significant portion of CHI's work for the Federal government in the 1990s - see Figure I-1.

The emphasis on outcomes was enforced by congress whose interest in performance and results was expressed in GPRA - the Government Performance and Results Act of 1993. Cozzens has analyzed the relationship between GPRA and research evaluation. The act calls on every Federal agency to do strategic planning, to report annually using performance indicators, and to use program evaluation in its strategic planning and performance reporting efforts. No agency is to be exempt, including research agencies. Full implementation is due shortly (Cozzens, 1995b). Cozzens writes that GPRA enshrined the distinction between outputs and outcomes and stated that outcomes are of primary interest. GPRA is problematic for research because it requires agencies set goals and measure their progress toward them. This makes sense for research activity (number of grants awarded, number of research students supported etc.), but not for the use and application of research which is largely out of an agency's control (Cozzens 1995a).
GPRA served to increase interest in non-peer review evaluation in another way. Cozzens 1995b writes:

> Most of the assessment of Federal research programs is descriptive, and far removed from the sort of quantification of performance the Government Performance and Results act requires. A large array of quantitative tools for evaluation has been described in the literature, but few of them are used in practice. To respond to GPRA, research programs and agencies thus face the challenge of either choosing among an array of options they have largely avoided in the past or developing new ones. (p. 354)

GPRA's requirement to quantify provoked renewed discussion of evaluation methods in Washington under the guise of which appeared thinly disguised reiteration of the value of peer review (COSEPUP, 1999). Cozzens (1995b) describes four agency pilot projects to develop acceptable indicators in an illustration of the underlying tension between agencies avoiding quantification and the Office of Management and Budget demanding measurable indicators.

Another factor may be at work to increase the use of non-peer review evaluation, that is an increased complexity in the nature of granting programs. The focus on outcomes has accompanied policy interest in science-industry links. This has led to policy and programs to try to increase linkage - such as Cooperative Research and Development Agreements (CRADA's) that provide a framework for national laboratories to work with industry with the goal of increasing technology transfer. Money for research centers with industry involvement required also became prominent.

Evaluations conducted in the late 1990s reflect this. The previously mentioned evaluations conducted by Abt Associates address the new complex programs (Small Business Innovation Research; Design, Manufacture, and Industrial Innovation Program; Science and Technology Centers Program; Engineering Research Centers). Mechanisms such as these, with goals that go beyond fostering research excellence in traditional scientific disciplines, may well need a more comprehensive and structured external evaluation than is provided by peer review. Traditional paper-paper bibliometrics alone tends also to be insufficient, which may explain a change in CHI's work. In the 1980s, CHI's paper-paper, or science bibliometrics, tended to stand alone. In the later half of the 1990s, CHI's paper-paper bibliometrics were combined with case study methodology in evaluations performed collaboratively. This movement to multiple methods suggests that more complex methods are required to evaluate more complex programs.

Evaluators and the evaluated

Fragmentation characterizes research evaluation in the United States. Melkers and Roessner (1997) state that:

> Among the industrialized democracies, the political system of the U.S. stands out because of its relatively fragmented distribution of authority, its responsiveness to organized interests at multiple points of access, and its heavy reliance on bargaining as a decision-making tool . . .. Rather than exhibiting a single or dominant focus for the support and use of program evaluation, the United States government is characterized by multiple locations in both the legislative and executive branches. Evaluation performers are even more widely scattered, with both branches of government possessing multiple in-house capabilities, and both contracting out to a variety of different types of performers that include universities, private companies, not-for-profit organizations, and quasi-public agencies such as the National Academy of Sciences. (Melkers & Roessner, 1997, p. 58)

The fragmentation in evaluation of Federal programs extends to evaluation of R&D programs and makes it difficult to be comprehensive in an overview. For a reviewer, the fragmentation originating in the political situation is compounded by the limited publication of evaluations. Availability is important here, because without a public record, reviews cannot be comprehensive. All evaluations should be available in theory because the Federal government adheres to a freedom of information principle. In practice, publication makes all the difference, and evaluations are not often published. Even if an agency will allow publication, evaluations are commissioned to address issues specific to part of an agency at a point in time. The lack of generality in this makes the material less than ideal for publication. A public record does not accumulate also because private firms conduct many evaluations and they are generally less interested in publication. Clearly the Internet has changed this, and evaluations such as Roessner et al. and the Research Value Mapping project summarized in this report are now easily available thanks to the Internet (although both authors are also academics who have published articles from this work).

A publication record is important because evaluations are ephemeral. Particular to a time and place, as time passes and the situation changes, interest in old evaluations evaporates and they sink from view rather quickly. When there is turnover in agency staff, and those who commissioned an evaluation leave, institutional memory can be lost. Thus, even agency staff cannot identify all the evaluations conducted by their agency. Only Logsdon and Rubin undertook a survey to determine the true situation, and even they were reduced to offering a one or two examples from an agency with the awareness that more had been conducted.

Evaluations would be less ephemeral if they had greater impact. For example, CHI's study of patent-paper linkage in the late 1990s established that most work cited in patents came from the public sector, and it is still remembered because it was the subject of an article in the New York

Times and was used by the Administration in arguments supporting science budgets. However, this was not strictly speaking a program evaluation, and so was more general in its implications. The lack of impact that evaluations have on decision making is often lamented, particularly in reviews of the early 1990s. Yet most evaluations seem to be commissioned to serve a particular need for information at a particular time and if the information produced is particularly favorable, may have a limited impact on a small circle concerned with the program investigated. It is difficult for performers to identify the impact of evaluations, because impact occurs in an extended process after the performer has moved on to other work.

With the exception of Kostoff's work, reviews of evaluation were associated with the Office of Technology Assessment (OTA) until its demise in the mid-1990s. The OTA wrote two reports addressing research evaluation, in 1985 and 1991, and they commissioned background papers from academics. Clusters of journal articles reviewing evaluation in the U.S. appeared around the time of the OTA reports. With the demise of the OTA, the situation changed and Rand produced a review in 1995, Melker and Roessner in 1997, and Roessner and Georghiou produced another in 2000 for a special issue of *Research Policy*. These reviews, with the addition of CHI's in-house records, form the basis for this overview.

As for the evaluated, several agencies stand out. The Office of Naval Research (ONR) performs basic research, and is the home of Ronald Kostoff, a prolific writer on the subject of research evaluation whose publications are on the internet and so easily available. Kostoff has performed research evaluations at ONR and has published some of them. His most recent large effort is summarized in Chapter 4. The National Science Foundation, also focused on basic research, commissions a fair amount of evaluations - for example Roessner et al.'s tracings and the work described on the Abt Associate web site. There seems to be activity at the Department of Energy, though it also seems particularly fragmented and difficult to know in its entirety.

In recent years, the ATP program has stood out from the crowd on several fronts, changing the evaluation landscape in the U.S. almost single handedly. Under heavy congressional scrutiny, it has performed repeated evaluations of itself and in turn been evaluated by the General Accounting Office (GAO). It has commissioned work from a number of consultants in private practice and academia, and this work has been published. This work has largely been economic; ATP employs economists on its staff; and the agency has pushed the development of economic methods as applied to evaluation.

The fragmentation in the research evaluation area makes it difficult to provide a complete list. No doubt some have been inadvertently left off the following list of performers of research evaluation:

- Abt Associates – Has conducted evaluations using case study methodology for the National Science Foundation.
- Prof. Harvey Averch, Florida International University
- Barry Bozeman et al. – Behind the Research Value Mapping project is Bozeman and colleagues at Georgia Institute of Technology who are examining evaluation from a perspective grounded in social science theory.
- CHI Research, Inc. – conducted bibliometric evaluations for various agencies for two decades.
- Prof. Susan Cozzens – An observer and commentator on the evaluation scene, particularly on GPRA and evaluation.
- Prof. Irwin Feller, The Pennsylvania State University

- General Accounting Office – Supports Congress and conducts evaluations at its request. GAO conducts the work itself. See Melkers & Roessner, 1997 for detailed analysis of several GAO studies.
- Ronald Kostoff – At the ONR. Employs bibliometric techniques and is developing data mining applications.
- The National Academies – National Academy of Sciences, Institute of Medicine, National Academy of Engineering, National Research Council and their committee: COSEPUP – Committee on Science, Engineering and Public Policy. The Academies serve the nation by producing a large number of reports advising on current issues involving science and technology. The reports are based on the work of panels of the nation's top scientists and technologists. In the area of research evaluation, the National Academies represent the interests of the scientific community and advocate and perform peer review evaluations.
- SRI International – Also a major performer of evaluations. David Roessner, also an academic at Georgia Institute of Technology has conducted some of their evaluations.

Those who have performed evaluations for the ATP program include:

- Drs. David Austin and Molly Maccauley, Resources for the Future, Washington D.C.
- Professors Lee Branstetter, UC-Davis
- CONSAD Research, Inc.
- Dr. Henry Etzkowitz, State University of New York at Purchase, NY
- Dr. Maryann Feldman, Johns Hopkins University
- Prof. Paul Gompers, Harvard University
- Prof. Zvi Griliches, now deceased, Harvard University
- Prof. Josh Lerner, Harvard University
- Professor Albert Link, University of North Carolina at Greensboro
- William F. Long, Business Performance Research Associates
- Haim Regev, Israeli Office of Statistics
- Research Triangle Institute
- Mariko Sakakibara, UCLA
- Solomon Associates, Silber & Associates - Two consulting firms in Washington D.C. that performed early surveys for ATP.
- Manuel Trajtenberg, Hebrew University, Israel
- Professor Nicholas Vonortas, George Washington University
- Professors Todd Watkins and Theodore Schlie, Lehigh University

Conclusion

In the United States, internal evaluations have predominated. Peer review, whether conducted by agency staff or by The National Academies, is the most widely used and generally accepted method of evaluation. External evaluation, using outsiders who bring methodological expertise, is said to be much less prevalent and is subject to much more reviewer criticism. The fragmented situation that surrounds evaluation makes it likely that reviews of evaluations, mostly written by academics, underestimate the frequency of evaluation, mostly performed by consulting companies. Whatever the situation was, it seems to be changing. The Internet reduces fragmentation by making unpublished work more widely available and keeping work available for longer. More external evaluation is being performed, and the forces at work are not abating, so this trend should continue.

Chapter II. Evaluation Approaches Used by the Advanced Technology Program, 1990-2000

by Rosalie Ruegg, Technology Impact Assessment (TIA)

The Advanced Technology Program (ATP) began development of its evaluation program soon after the program began operating in 1990. One factor that stimulated the early start at evaluation was a legislative requirement that the ATP provide an assessment four years after the program began. Another was a view by program managers that evaluation is an important feature of good management. Passage of the Government Performance and Results Act (GPRA) in 1994, and constant external requests for information provided additional impetus to evaluate.

The ATP's evaluation program was developed and is led by its Office of Economic Assessment, with considerable input from outside experts in the field. Academic and consulting economists and other evaluators contributed importantly by participating in planning workshops and by carrying out evaluation studies. Staff of the Economic Assessment Studies have coordinated the overall evaluation program, carried out research, managed contracts, built databases to track progress, presented findings to program administrators and policy makers, and provided feedback from evaluation to project/program selection.[2]

At the end of its first decade, ATP's evaluation program is receiving high marks. A recent independent assessment by the National Academy of Science of ATP and other partnership programs concluded, "the ATP assessment program clearly surpasses other U.S. partnership programs with its rigor, scope, and independence." Dr. Irwin Feller, a leading economist in the field of evaluation, described ATP's evaluation program as "a model for other partnership activities."[3] Even the Office of Inspector General of the U.S. Department of Commerce, charged with finding waste, fraud, and abuse in government programs, had praise for ATP's evaluation program, calling it "exemplary."[4]

An eclectic approach to evaluation

In the spirit of using the best tool for the job, the ATP has used a variety of evaluation methods, both qualitative and quantitative, to assess the effectiveness of ATP awards.[5] While this approach may have initially seemed somewhat fragmented, the passage of time has helped

[2] In addition, the Economic Assessment Office staff, like other members of ATP, served on project selection boards and project management teams.

[3] National Academies of Science, National Research Council, Board on Science, Technology, and Economic Policy, *Government-Industry Partnerships: The Advanced Technology Program; An Assessment of the Advanced Technology Program* (Washington, D.C.: National Academy Press, 2001),

[4] U.S. Department of Commerce, Office of Inspector General, Report on the ATP.

[5] Throughout the program's existence, it has used peer evaluation to select projects to fund, sometimes informed by technology and market assessments. It has also relied, like most government programs, on regular assessments by advisory boards and oversight committees of its processes, operations, and decision-making strategies. Those form of evaluation are not the subject of this report. Rather, the focus here is on evaluation of projects funded.

reveal the larger fabric of the evaluation plan and how the pieces fit together. Some of the studies have focused on ATP's effect on the innovation process; some on the commercial progress of the innovators; some on spillover effects; some on the workings of collaborative research; some on improving the tools of evaluation; and others on combinations of effects. Some of the studies have been extensive; others modest in scope.

The principal objectives of ATP's evaluation program have been, first, to increase understanding about relationships among various program inputs, outputs, and outcomes in order to improve the partnership program; second, to track and measure the performance of individual projects in order to know what is working and what is not; and, third, to measure the order of magnitude of net social benefits from the program's portfolio of projects, with attention to private benefits, spillover benefits, societal benefits at large, and that part attributable to the ATP, to assess whether the program is meeting its goals.

Here, a sampling of methods used by the ATP for evaluation is presented in the context of the studies in which they appear, and also in the historical context of the evaluation program's evolution. References to the full studies facilitate reader access to additional information. The focus in on published studies, although several studies in progress are also mentioned.

Because the ATP is funding applied research for technology development, the evaluation program has been able to push towards the more quantitative, results-oriented, economic methods of assessment.[6] Principal evaluation methods used by the ATP have been statistical profiling to track the make-up of the portfolio of projects; survey of funding recipients and non-recipients; case study, including anecdotal studies with and without systematic collection of output data; case study using the benefit-cost method, expected value analysis and sensitivity analysis; and a variety of statistical and econometric methods, including production functions, regression analysis, cost-index models, and macroeconomic models. To a lesser extent, bibliometrics and patent citation analysis have been used for ATP's evaluation.

Statistical characterization of portfolio attributes

The first priority for evaluation was to be able to characterize the most important dimensions of the newly unfolding program: the applicants, the awards, the projects, the participants, and the technologies funded. This information was needed for program management purposes; it was also needed for evaluation. How successful was the program in attracting businesses to apply? To what extent was the program attracting joint ventures versus single-applicant projects? How many awards was the program making? How much were ATP and its partners putting into the projects? How many companies, universities, and other organizations were involved in the projects? Which organizations were involved? Where were they located? Were small businesses able to compete? What technology areas were funded? How long were the projects lasting? How many awards were going to foreign-owned companies? How many projects had completed? How many were still underway? How many were stopped prior to completion? Why? These are examples of the many questions asked of the ATP by members of Congress and others.

Constructing and maintaining databases for providing quick responses to questions such as

[6] In contrast, Federal programs that fund mainly basic research usually find economic and financial methods more difficult and problematic to apply.

those listed above was the first-order of business. Now, collecting and reporting these data are routine tasks for the ATP. The core databases of applicants, awardees, and technologies are also the starting informational point for most evaluation studies.

Survey method

The ATP has used the survey method extensively, gathering information primarily from applicants and awardees about their perceptions and experiences with the program, how they perceive value, and how they are affected. An advantage of the survey method is that it provides a relatively quick way to obtain qualitative and quantitative information about program effectiveness that is easily understood by a broad audience. But care must be taken to design survey instruments to avoid biasing results, and the response rate may be low. Also those surveyed may require anonymity in order to answer honestly, and this requirement may restrict handling and use of the data collected.

Solomon Survey: In order to gain early information about project performance, ATP commissioned a survey of the first group of award recipients, near the end of the first year of funding. Solomon Associates, a small consulting firm in Washington, DC, carried out this survey. [7] The survey was conducted by telephone interview, with the use of a number of open-ended questions to allow the respondents latitude to identify the areas that seemed important to them. The responses were then coded and tabulated. The study reported that the awards had accelerated research, suggested the importance of collaboration, indicated improved competitiveness, and identified a "halo effect" indicating that award recipients were being successful in attracting additional funding from the private section. These and other findings were reported as short-run effects.

Silber Survey: The "Solomon Survey" was followed several years later by a larger survey, the "Silber Survey," conducted by a small consulting firm located in Maryland.[8] The second survey covered all award recipients funded in the first three years of the ATP. The Silber Survey used more closed-end questions than the earlier survey and the questions were informed by the results of the earlier study. ATP economists worked closely with the consultant to design the study. It was comprised of two parts: one focused on early effects of the award, the other on "customer satisfaction," that is, how companies found it to work with the ATP through various stages of the partnership. Negative as well as positive results were reported, with respondents guaranteed anonymity and the database maintained by the contractor. The survey identified such effects as the difference the ATP was thought to have made to the participants in terms of their ability to afford and engage in high-risk research; their likelihood of pursuing the proposed technology development projects with the same goals, level of effort, and speed without ATP funding; corporate research investment stimulated by the award; the extent of collaboration and its positive and negative effects on award recipients; effect of the award on commercializing resulting technologies; attraction of follow-on funding; changes in international competitive

[7] Solomon Associates, *Advanced Technology Program: An Assessment of Short-Term Impacts -- First Competition Participants*, NIST Contractor Report, February 1993. The report is available only in hard copy from the ATP Office of Economic Assessment.

[8] Silber & Associates, *Survey of Advanced Technology Program 1990-1992 Awardees: Company Opinion About the ATP and its Early Effects, NIST Contractor Report*, January 1996. The report is available on-line (www.atp.nist.gov/eao/eao_pubs.htm).

standing; changes in company employment; changes in the way of doing business; willingness to share non-proprietary information; and other effects.

A copy of the survey instrument used for the Silber Survey is included in an appendix to the Silber study report. There are two versions of the questionnaire: a "long form" administered to all single applicants and to those joint venture participants (JVP) who were identified as most likely to be involved in subsequent commercialization activities; and a "short form" administered to the JVPs who were identified as less likely to be involved in commercialization efforts. The difference between the two versions was a 48-question section on commercialization and business goals.

ATP's Business Reporting System: ATP's assessment staff, realizing that it needed the ability to track these and other effects on an on-going basis, developed in-house capability to collect data directly and electronically from project participants. Called the "Business Reporting System" (BRS), this in-house tracking system enabled the ATP to report on program developments on a regular basis. For all projects funded after 1992, the BRS tracks project progress during the period of ATP funding, and then follows further developments in the post-project period.[9] The individual company reports are held as confidential and proprietary by the ATP, because they contain certain business planning data that some of the companies wish not released. Only aggregate reports of the BRS data are published.[10]
The BRS consists of several parts. At the beginning of the project, participants report their planned application areas for the technology and strategies for commercialization. Annually they report on progress towards implementing their commercialization strategies and on short-term economic impacts of the projects, such as early sales revenues, shortened R&D cycles, collaboration effects, intellectual property creation and protection, and early job creation. At the end of the project, they report on accomplishments and future plans. In the post-ATP funding period, they report on accomplishments towards commercializing and diffusing the technology. They are also asked to identify any other organizations that are pursuing commercialization of the technology. Over the six-year period following project completion, participants report three times, increasingly emphasizing economic impacts to the nation, provided they are making commercial progress.

Cycle-Time Survey: In addition to these broad-coverage surveys, there have been more narrowly focused surveys conducted to answers specific questions or test hypotheses. One of these was a study conducted by an ATP industry consultant, Dr. Francis Laidlaw, to investigate the impact of ATP participation on the applied research cycle time of program participants.[11] Cycle-time impacts are of keen interest because of the program's mandated emphasis on speeding the creation and application of advanced, generic technologies.

---

[9] Jeanne Powell in "The ATP's Business Reporting System: A Tool for Economic Evaluation," Paper presented at Conference on Comparative Analysis of Enterprise Data, Helsinki, Finland, 17-19, June 1996, describes the BRS.

[10] Examples of reports based on BRS data, starting with more recent reports, are the following: Jeanne W. Powell and Karen L. Lellock, *Development, Commercialization, and Diffusion of Enabling Technologies: Progress Report*, NISTIR, April 2000; Jeanne Powell**,** *Business Planning and Progress of Small Business Firms Engaged in Technology Development through the Advanced Technology Program,* NISTIR, October 1999; and Jeanne Powell, *Development, Commercialization, and Diffusion of Enabling Technologies: Progress Report for Projects Funded 1993-1995*, NISTIR, December 1997.

[11] Francis Jean Laidlaw, *Acceleration of Technology Development by the Advanced Technology Program: The Experience of 28 Projects Funded in 1991*, NISTIR, October 23, 1997.

Earlier surveys had found that a majority of program participants believed that participation in the ATP had helped them to reduce cycle time and that it was important to do so - but the earlier surveys did not provide details on why cycle-time reduction is of special importance to the companies, or how participation in ATP helped them to reduce cycle time, whether reductions in R&D cycle time were carrying over to reduce commercialization time and get to market more quickly, or whether there were temporal effects extending beyond the ATP project. Laidlaw conducted structured telephone interviews with the primary investigators of 28 projects funded by the ATP in 1991. The majority indicated that participation in ATP resulted in cycle-time improvements both in R&D and in commercialization, and that the improvements carried over to other technology development projects outside of ATP. They spoke of adapting specific "ATP practices" to related projects; application of methodologies and processes used or developed in the ATP project to the firm as a whole; development of a cultural bias favoring speedier processes; and accelerating the development of a whole series of applications from the ATP-funded technology platform.

Feldman/Kelley Survey: An important, recent study that also used the survey method is one led by Dr. Maryann Feldman of Johns Hopkins University, in collaboration with Dr. Maryellen Kelley then of the ATP's economic staff.[12] The purpose of the survey was to collect information on the preparation of ATP proposals, the involvement of other organizations (formally or informally) in ATP proposed projects, the views of applicants concerning the fairness of the selection process, and post-competition experiences of winners and non-winners of awards. A point of particular interest was whether non-winners had been able to proceed with their proposed research plans, and whether winners and non-winners differed in their subsequent ability to attract other sources of funding. The use of a control group of non-winners strengthened the reliability of the results. (Previous surveys had employed an implied counterfactual question of award winner to get at this question, e.g., without the ATP award, would they have behaved differently, and how.)

The survey started with 502 proposals submitted in the 1998 ATP competition. There were a total of 822 organizations represented in this competition. Interest centered on 741 for-profit enterprises that applied to ATP in 1998, and this group of firms constituted the sampling frame. The sample selected consisted of 100% of the winning firms and a simple random sample of 50% of the non-winners. The survey instrument was pre-tested, and also the Office of Management and Budget reviewed it in conjunction with requirements of the U.S. Paperwork Reduction Act.

All interviews were completed over a six-month period (June-December 1999). The analysts followed standard survey method procedures. Respondents were promised confidentiality and assurance that steps were taken to ensure that their responses to any of the survey questions would remain anonymous, and would not be publicly released in any form that would identify a specific individual or company. In addition, a letter from the contractor to those to be interviewed identified the organizations responsible for carrying out the survey (Johns Hopkins University and the University of Baltimore) and included a selection of questions that the respondent might find helpful to have in advance of the telephone interview.

[12] Maryann Feldman and Maryellen Kelley, *Leveraging Research and Development: The Impact of the Advanced Technology Program*, NIST report in publication. The report is expected to be on line in the spring of 2001 (www.atp.nist.gov/eao/eao_pubs.htm).

Outcome indicators

ATP compiles a series of indicators of outcomes, some of which are used for budgetary and GPRA reporting. These include funds available annually for project awards (an input measure), number of projects funded each year, cumulative number of projects funded, cumulative funds awarded, number of project participants by type, number of projects completed, number of technologies under commercialization, number of patents filed and granted, number of publications, as well as others. These indicators are drawn primarily from the BRS database.

Case study method[13]

The ATP has found considerable interest among stakeholders in the "story" of technology development, including the human side, the research, the technical challenges, and resulting applications. Each project funded has a unique story, and the case-study method offers the advantage of telling the story of a project or set of projects. Some case studies are purely anecdotal, designed to promote understanding and appreciation of the program and how it works, the science and technology, and the potential value of providing public support to research. Closely related are case studies that trace the evolution of a field of research. Others link project descriptions to quantitative analyses, such as statistical or benefit-cost analysis. Some case studies provide brief, snapshot treatments of projects. Others provide in-depth analysis. In short, the case study method has been much used by the ATP, in a variety of ways. Examples of six difference types of case studies commissioned by the ATP follow:

<u>Anecdotal Case Studies</u>: Over the years, a number of anecdotal case studies of projects have been prepared. An example is provided by Professors Paul Gompers and Josh Lerner of Harvard University's Business School who performed seven anecdotal case studies of ATP-funded projects as part of a larger statistical study on venture capital funding.[14] Their objective for the case studies was to learn why the companies needed government funding and to understand conditions at the companies immediately before and after ATP funding was obtained. They conducted largely unstructured interviews with seven start-up biotechnology companies all located in the Boston area. The stories center on the difficulty of high-tech, start-up companies in obtaining funding to pursue technology development that is considered high risk. The stories refute the idea that the growing amount of venture capital in the U.S. has made it easy for small, high-tech companies to obtain funding for the kinds of high-risk, enabling technology targeted by the ATP.

<u>Genesis Case Studies</u>: From 1994 through 1998, the ATP awarded most of its funding through focused programs, each with a unifying set of project goals. Thirty focused program competitions were organized and held by the ATP to fund suites of projects to address pre-defined problems. This approach was controversial, with considerable misconception and criticism concerning how the focused areas were selected.

---

[13] The typology of evaluation methods is not standardized. Alternative groupings of methods are possible.

[14] The seven cases are contained in the following report that focuses on venture capital: Paul Gompers and Josh Lerner, *Capital Formation and Investment in Venture Markets: Implications for the Advanced Technology Program*, NIST Contractor Report, December 1999.

To inform public policy, Dr. Henry Etzkowitz, a professor of sociology at the State of University of New York at Purchase, NY, together with Dr. Richard Spivak, economist on ATP's staff, conducted a case study to trace the genesis of one of the ATP's focused program – the Information Infrastructure for Healthcare Focused Program (IIH).[15]

The authors recount how company researchers, managers, academics, association members, and private citizens responded to ATP's "White Paper Process" by submitting many ideas for research, a significant number of them in the healthcare area. The authors describe how the private sector's response in the area of IIH differed from ATP's expectations, by generating many fragmented ideas rather than well-developed program ideas. Yet, taken together, the white papers formed the basis for sketching out a potentially comprehensive focused program. The authors then describe subsequent steps leading to development of the IIH Focused Program: the development by ATP staff of a comprehensive White Paper from the collection of individual papers; public workshops to make more widely known the shaping of the focused program and to solicit additional input from a wider audience; the internal NIST review process; and additional meetings to publicize the IIH Focused Program. A follow-on study by Ms. Bettijoyce Lide and Dr. Richard Spivak, ATP project manager and business specialist, respectively, for the IIH Focused Program, continued the unfolding story of the focused program by describing projects funded and future opportunities within its umbrella.[16]

<u>Set of Anecdotal Case Studies with Systematic Collection of Output Data</u>: By 1997 the ATP had completed a number of in-depth case studies, but was looking for an approach that would provide systematic coverage and avoid the issue of case-selection bias. The approach adopted was to perform anecdotal studies for all completed projects approximately two years after their completion, and supplement the anecdotal approach with systematic collection of selected output data and aggregate statistical analysis. In addition, results of detailed case studies were brought into the project presentation when they were available.

The first report on performance of completed projects was published in 1999.[17] It covered the first 38 projects.

The most recent report on performance of completed projects covers 50 projects. It subsumes the previous 38 project write-ups, and adds 12 new completed projects, patent tree analysis, composite performance scores, and a new analysis of aggregate performance.[18] It also adds additional analysis to the report of projects stopped prior to completion, contained in both the earlier and later reports.

---

[15] H. Etzkowitz and R. Spivak, *Information Infrastructure for Healthcare: An Evaluation of a Government-Industry Technology Development Initiative*, NIST Contractor Report, October 1999.

[16] B. Lide and R. Spivak, *Advanced Technology Program Information Infrastructure for Healthcare Focused Program: A Brief History*, NISTIR, February 2000.

[17] William F. Long of Business Performance Research Associates (Author), (Rosalie Ruegg, ATP Project Manager), *Performance of Completed Projects*, Status Report Number 1, NIST SP950-1, March 1999. The report is available at ATP's website (www.atp.nist.gov/eao/eao_pubs.htm).

[18] Rosalie T. Ruegg of TIA Consulting (Senior Author), (Darin Boville, ATP Project Manager), *Performance of the First 50 Completed Projects*, Status Report Number 2, NIST SP950-2, publication expected in March 2001, to be followed by addition of the report to ATP's website (www.atp.nist.gov/eao/eao_pubs.htm).

The performance ratings for the 50 projects are based on progress in both creating and disseminating new technical knowledge to others, as well as the direct use of the new knowledge by the innovators to accelerate commercial use of the technology – three dimensions of performance that figure prominently in achieving the long-run success of the ATP. Slightly more of the 50 projects are rated weak performers rather than strong, and most of the project fall somewhere in the middle. Yet, expected net benefits from the strong performers alone are more than enough to yield a robust performance for the group of 50 taken as a whole. Aggregate estimated benefits for the strong performers, in fact, outweigh total ATP costs to date. The following caveat applies: The projects were assessed in two different groups at two different calendar times – 38 of the projects were assessed two years ago and 12 more recently. Since technology development and commercialization take time and are characterized by unexpected breakthroughs and failures, future updates of these projects may alter the reported findings.

A major advantage of the approach is that the work is considered particularly significant because it shows what has happened with a large group of ATP-funded projects several years after completion, and it avoids the selection bias that reduces confidence in the selection of individual projects for case study. Another advantage is that the resulting report is appealing to a large, diverse audience, ranging from policy makers to investors.

Disadvantages are that, given the time and cost of conducting detailed case studies, the approach to evaluating completed projects has been to limit the depth of investigation as the number of projects to be evaluated increases. Further, it is difficult to perform and assemble a large number of cases and their overview in a short period of time; hence, the earlier-done cases tend to become dated while the later ones are being prepared. The cases provide a snapshot of project developments; the final project outcome may differ from the suggested outlook.

In-depth Case Study Measuring Research Cost Savings from Collaboration Efficiencies:
ATP's first in-depth project case study focused on a joint venture project and was initiated when the five-year research project had been underway only two years. A second, follow-on study was conducted by the same researcher at the project's end.

The joint-venture project aimed at developing a suite of "leap-frog" technologies for the U.S. printed wiring board industry, which at the time was in decline. Users of the boards had indicated that they would have to turn increasingly to overseas producers if the domestic suppliers were not able to make major technological advances within the next five years. With encouragement and administrative leadership from the National Center of Manufacturing Science (NCMS), a group of seven research-capable companies within the industry, plus Sandia National Laboratories, carried out the research program to increase competitiveness of U.S. producers. The project proposers identified a large number of potential innovations and developed a research agenda around them. Although the proposing organizations stood to benefit directly either as producer or user of the wiring boards, the project also included a user group intended to improve the international competitiveness of the entire U.S. industry sector. The project lasted from mid-1991 through mid-1996.

Conducted by Professor Albert Link of the University of North Carolina at Greensboro, both studies focused on the impact of the collaborative effort on research costs and timing of innovations. To a lesser extent, the second study also investigated early-stage benefits to companies from adopting the new technologies, but since it was conducted at the end of the study, there was as yet only limited adoption of the innovations.

Link used a survey instrument to collect data from the participating organizations on their research plans, estimated costs, competitive positions, and other factors. The responses were used to estimate the impact of the participants' collaborative activities on R&D efficiency.

In order to establish a lower bounds estimate of the cost savings attributable to the ATP, the participating companies were asked to separate out those research activities they would eventually have done without the ATP from those that they would not have done at all without the ATP. This constitutes the study's counterfactual analysis. Then, they were asked to estimate how much it would have cost them to do the set of activities they said they would have undertaken without the ATP, assuming the ATP project had not occurred. The estimated cost without the ATP, minus the cost with the ATP for the set of activities they would have undertaken anyway was taken as a minimum value of the ATP award. The estimate was that undertaking the research as a collaborative project yielded a cost savings of approximately $35.5 million. Not counted in the estimate was the half of research activities that the companies said they would otherwise not have undertaken at all, since the concept of cost savings does not apply in that case. Also not included in the estimate was the fact the participants said the ATP project accelerated the research they otherwise would have eventually undertaken.

The increased research efficiency was estimated in turn to have reduced cycle times for both new project development and new process development. The research effort gave the industry new technical capabilities. The project, it was concluded, has meant productivity improvements for member companies, diffusion of the technology to other producers, and improved competitive positions of U.S. producers in world market.

In-depth Case Studies Using Benefit-Cost, Expected Value, and Sensitivity Analyses:
To date, economists at Research Triangle Institute's Center for Economic Research conducted the most extensive set of in-depth case studies for ATP using benefit-cost analysis.[19] The study developed a case-study evaluation framework for assessing medical technologies and applied it to seven tissue-engineering projects funded by the ATP between 1990 and 1996. Included in the study were new technologies for the diagnosis and treatment of cancer; the treatment of diabetes, damaged ligaments, tendons, and articular cartilage; and transplanting xenogeneic organs.

The methodology takes a very detailed and specific look at each technology project, analyzing how the technology works and how and when it might be used, what effect the ATP has had on its development, the past and future costs incurred in its development and commercialization, and its value in the expected first application. That is, the study characterizes the entire R&D-commercialization-production-health process for each technology.

In terms of benefits, the model accounts for reductions in direct costs of medical treatment, and, where supporting data are available, reductions in the costs of patient pain and suffering. The economic burden of illness and disease potentially includes (1) direct medical costs in the form of explicit payments for prevention and treatment; (2) indirect costs in terms of productivity losses and the implicit value of the resources expended by uncompensated caregivers; and (3) intangible costs, comprising the pain and suffering incurred by patients and patient' families and friends. RTI's model includes (1) direct medical costs and (3) the intangible costs, but excludes (2) indirect effects.

Central to the study's evaluation of benefits from improvements in a patient's well being are two concepts: Quality-Adjusted Life Year (QALY) and value of life. QALY is an approach to

---

[19] Research Triangle Institute, *A Framework for Estimating the National Economic Benefits of ATP Funding of Medical Technologies*, April 1998. Report is available on-line (www.atp.nist.gov/eao/eao_pubs/htm).

quantifying health benefits to an individual in terms of the quantity and quality of life.[20] A year of life in full health is given a QALY value of 1.0; death is given a QALY value of 0.0; and a year of life at less than full health is given a QALY value between 0.0 and 1.0. (In fact, some conditions are assigned a negative value indicating they are considered worse than death.) QALY values for selected health states are reported in health assessment literature. For example, living with mild angina has been assigned a QALY value of 0.90, and severe angina, a QALY of 0.50.[21] Thus, living a year with severe angina is valued the same as living a half a year in a healthy state.

QALY values assigned to different health states are derived from averages of survey results for relevant populations. QALY values allow the analyst to quantify health improvements by accounting for changes in quantity and quality of life in a single measure.

"Value of life" as used in the study means the value of a statistical life as indicated by a collective willingness to pay to avoid fatality risks.[22] The value of a life year in perfect health can be derived from the value of a statistical life. Changes in QALY associated with a given medical technology can be applied to the value of a life year in perfect health to generate an estimate of a patient's health benefit.[23]

The framework accounts for both chronic and acute illness and injury models: (1) The progression of chronic diseases is modeled as a Markow process where patients transition from one health state to the next throughout the statistical remainder of their lives. (2) Acute illness and injury is modeled as a single-period case of the chronic disease model.

A Bass diffusion model[24] accounts for the process of technology diffusion. Interviews with experts about the likely application of each technology assist the analyst to estimate the rate and ultimate market penetration of each tissue-engineering project.

RTI analysts measure the economic benefits from the public and private investments in medical technologies in terms of net present value (NPV), i.e., dollars adjusted for the time value of money. Benefits and costs that occur in future periods are "discounted" to make them comparable with those occurring in the present.
The study identifies three ways in which ATP funding may increase social returns: (1) it can accelerate R&D, thereby leading to earlier introduction of the new technology. Receiving the benefits earlier or over more years will increase a project's NPV, other things being equal. (2) It

---

[20] George W. Torrance and David Feeny, "Utilities and Quality-Adjusted Life Years," *International Journal of Technology Assessment in Health Care*, 5, 1989, pages 559-575.

[21] *Ibid*.

[22] See Josephine A. Mauskopf and Michael T. French, "Estimating the Value of Avoiding Morbidity and Mortality from Foodborne Illnesses," *Risk Analysis* 11(4), 1991, pp. 619-631; and Michael J. Moore and W. Kip Viscusi, "The Quantity-Adjusted Value of Life," *Economic Inquiry* 26(3), 1988, pp. 369-388.

[23] For additional description of the use of QALY values and value of life data, see Andrew Wang, "Key Concepts in Evaluating Outcomes of ATP Funding of Medical Technologies," *Journal of Technology Transfer*, Vol. 23(2), Special Issue Editor: Rosalie Ruegg, summer 1998, pp. 61-65.

[24] See a description of the model in Frank M. Bass, "A New Product Growth Model for Consumer Durables," *Management Science*, 15(5), pages 215-227, 1969.

can increase the intensity of R&D, thereby increasing the probability of R&D success and, hence, expected NPV. (3) It can broaden the scope of R&D to include a wider range of potential applications and, hence, increased NPV.

The model provides for computation of three economic performance measures for each project: (1) social return on investment, (2) private return on investment, a component of social return, and (3) social return on public investment, the return on ATP's investment based on the difference in social return with and without the ATP. Each measure is computed in terms of NPV and in terms of a percentage rate of return. The analysts perform sensitivity analysis to reflect the large uncertainties in project outcomes. They calculate multiple NPVs and percentage rates of return as they vary input values.

To estimate the benefits of the new medical treatment technologies, "defender treatment technologies," i.e., the expected best alternative medical treatments that would be used if the ATP-funded technologies were not available, are identified for comparison. Differences in projected declines in health states associated with the new treatment technology versus the defender treatment technology are used to estimate patient benefits. For example, three diseases were identified as primary health complications of diabetes – retinopathy, nephropathy, and neuropathy – each leading to possible declines in health states, characterized by such problems as blindness, renal disease, and lower extremity amputation. If the new treatment technology and the defender technology result in different probabilities that the health problems will occur, this provides a basis for estimating the expected value benefits of the new technology.

The RTI approach also incorporates a counterfactual scenario to model the situation without ATP funding, for comparison with the situation with ATP funding. For the set of medical technologies, the chief effect of the ATP is to accelerate the development of the new technologies. Hence, to estimate the part of benefits specifically attributed to the ATP, the difference with the defender technology is taken only over the interval that ATP funding has accelerated development of the new technology. The purpose of the counterfactual analysis is to identify that part of benefits specifically attributable to the ATP funding.

Principal findings of the RTI study are that the projects have expected social returns much larger than their private returns, primarily due to projected positive spillovers to patients treated with the new technologies. The study concluded that the ATP played a significant role in increasing the expected returns on these projects to the developers and to society at large by accelerating the R&D phase of the projects and improving the probability of technical success. The estimated composite social return on ATP's investment in the seven projects is in excess of $34 billion, in net present value.

The study has major limitations, most of which are attributable to the scope of the model in combination with the early status of the projects to which the model was applied. Modeling the entire process from R&D to health outcomes for these projects that were still in an early stage of development at the time of the study required the use of a large amount of estimated data and the use of numerous assumptions because much of the data were lacking. As a result, the study's findings should be regarded as preliminary. This limitation has more to do with the projects evaluated than the model itself, and would be reduced if the model were applied ex post to assess medical technologies that has actually reached the medical application stage. Another limitation that has to do with the model itself is the modeling of the defender technology in a static mode. Despite the limitations, RTI's approach provides a useful framework for evaluating ATP's expected contributions to social welfare.

Expanded Case-Study Model with New Features: The ATP has commissioned several new case studies using benefit-cost analysis. One of these, by Professors Todd Watkins and Theodore Schlie, both of Lehigh University, proposes to apply state-of-the-art methods to estimate net social returns to a cluster of photonics and opto-electronics projects funded by the ATP. One of the proposed advances in modeling is the use of a control group of projects not funded by the ATP to estimate what would have happened without ATP funding; this instead of a hypothetical scenario. Another proposed advance is to explicitly track both market and knowledge spillovers into the network of customers, competitors, suppliers, and to others outside the innovators' markets using a combination of snowball interviewing process coupled with patent and publication citation analysis.

An additional proposed feature is a dynamic interpretation of displaced defender technologies in which the defender technologies will change in the comparison against the ATP-funded technologies. It is too early to determine whether the actual study will deliver the proposed methodological advances.

In-depth Case Studies Using Macroeconomic Modeling to Project National Benefits: In general, the ATP has opposed using macroeconomic models for estimating program outcomes. Given the small size of the ATP's budget and the huge size of the U.S. economy, any effect would generally be indeterminate from statistical error.

The ATP has, however, seen a limited opportunity to use macroeconomic modeling to estimate regional and national impacts of projects under certain conditions. The ATP has commissioned use of the REMI model (Regional Economic Modeling, Inc.) for two case studies. But both were conducted only after a microeconomic analysis determined that project effects can be reflected in REMI's input-output matrix, and that the project participants comprise a relatively large part of the industry sector.

Underlying the REMI model is a set of structural equations linking economic variables via theoretical and empirical relationships.[25] These relationships are parameterized with publicly available historical data. The model is implemented in a computer program. Use of the model allows regional or national economic effects to be estimated.

CONSAD Research, Inc. first performed a REMI analysis of an auto consortium project aimed at reducing dimensional variation in vehicles.[26] This study was a modest effort that projected future benefits for the project while it was still underway.

Dr. Mark Ehlen, NIST economist, recently performed a detailed case study that used the REMI Model. He assessed the economic impact of a new flow-control machining technology applied in the automotive sector.[27] First he estimated the changes in automobile industry output and prices caused by implementation of the new process technology. Then he used the REMI model

[25] A comprehensive description of the REMI model is provided by G. Treyz, D. Rickman, and G. Shao, "The REMI Economic-Demographic Forecasting and Simulation Model," *International Regional Science Review*, Vol. 14, No. 3, 1992, pp. 221-253.

[26] The "CONSAD Report" is not available on-line, but is available directly from ATP's Office of Economic Assessment (Request by calling 301-975-3589 or e-mailing janet.brumby@nist.gov).

[27] Mark Ehlen, *Economic Impacts of Flow-Control Machining Technology: Early Applications in the Automobile Industry*, October 1999. The report is available on-line (www.atp.nist.gov/eao/eao_pubs.htm).

to estimate the total effect on national output, employment, and personal income. He performed analyses for two scenarios: one that assumed limited take-up of the technology; the other which assumed more extensive take-up.

Statistical and econometric methods

The ATP has used a variety of statistical and econometric methods in its evaluation. Some principal examples are given below:

A New Model for Estimating Future Consumer Benefits from Innovation: Drs. David Austin and Molly Macauley, both economists at Resources for the Future in Washington, DC, drew on previous work by another economist to develop a new method for estimating consumer benefits from advanced technology.[28] The new method takes into account changes in the quality of service provided by the new technology.

Bresnahan[29] had increased the feasibility of estimating consumer benefits from quality of service improvements, heretofore, a problematic undertaking at best. He developed a cost-of-living index approach that makes it possible to compare observed price and performance for an innovated product against hypothetical, best available price and performance data assuming the technical advance did not occur. A problem with Bresnahan's model was that it aimed at retrospective evaluation, i.e. at estimating consumer benefits from already existing innovations. It was less suitable for forecasting consumer benefits from innovations still under development, such as ATP-funded projects.

Austin and Macauley extended Bresnahan's method to make it suitable for prospective assessments, i.e., for estimating consumer benefits from proposed R&D projects and those that have not yet resulted in products on the market. Their approach allows for the gradual diffusion of the new technology. The model's parameters are expressed as probability density functions to reflect uncertainties over future or estimated parameter values. The two researchers also extended Bresnahan's method so that it would reflect consumer's preferences for specific product characteristics, such as speed, and to take into account the fact that those preferences may affect the product's success in a competitive marketplace.

Figure II-1 illustrates the expected gain in consumer benefits from a technological innovation, such as an improvement in the ability to store large amounts of data. The left-hand panel of the figure shows the pre-innovation baseline, where only a *defender technology* is available, whose supply is represented by S_0^{DT}. The downward sloping line, *D*, represents demand for data storage. The right-hand panel of Figure II-1 shows the ATP-sponsored innovation – a combination of cost reductions and quality improvements – that occurs in a subsequent time period. The innovation is represented as an outward shift in the supply curve to S_1^{ATP}. But in the meantime, the defender technology has also improved, represented by the baseline supply curve shifting out to S_1^{DT}. The shaded area represents the consumer welfare gain ("consumer

[28] D. Austin and M. Macauley, *Estimating Future Consumer Benefits from ATP-Funded Innovation: The Case of Digital Data Storage*, NIST Contractor Report GCR 00-790, April 2000. Available on-line (www.atp.gov/eao/eao_pubs.htm).

[29] Tim Bresnahan, "Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services," *American Economic Review*, vol. 76, no. 4, 1986, pp. 742-755.

surplus") at a point in time, due to the innovation. It is measured with respect to the hypothetical, future S_1^{DT} curve rather than the observed S_0^{DT}. As long as S_1^{ATP} lies to the right of S_1^{DT}, the innovation offers an improvement over the defender technology.

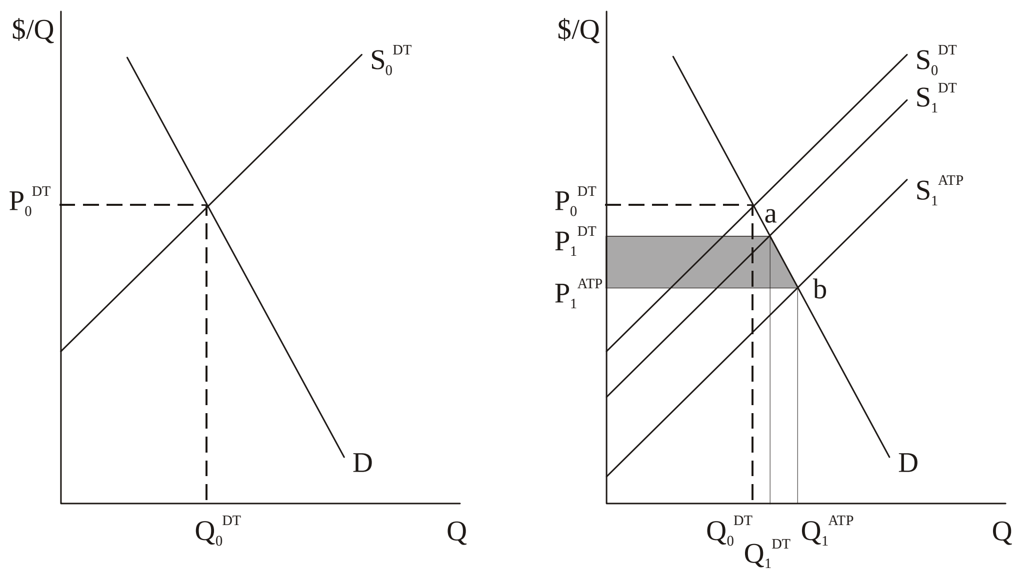


Figure II-1. Derived Demand for New Technologies: Illustration of Net Surplus Change

Measuring the gain is straightforward if the demand curve can be estimated using econometric techniques. But this is difficult to do in service sectors, where real output is not readily observed, yet where much of the demand for high technology is located. In this case, the Tornqvist cost-index approach pioneered by Bresnahan is attractive, because it does not require estimating a demand curve. The researchers paraphrase Brasnahan, stating that the method substitutes economic theory for (unobservable) data. With reference to Figure II-1, the cost index will be greater than unity, meaning costs are higher under the baseline scenario and consumers will be better off (gross of R&D costs) if the innovation occurs.

The index is an estimate of the change in the cost of living under the innovation scenario, relative to the baseline. To construct the index, the researchers adjust nominal unit prices of existing products or services to reflect consumer preference for improved quality/performance attributes.[30] They assumed that these shadow values decline over time, reflecting consumers' declining marginal utilities.

In their applications of the technique, they also adjust the prices of the defender technologies. The usually lower performance attributes of the defender technologies impose real user costs relative to the innovations. The price adjustments reflect consumers' willingness to pay to

[30] For a discussion of quality-adjustment methods employed by the Bureau of Labor Statistics in the construction of the consumer price index (CPI) can be found in B. R. Moulton and K. E. Moses, "Addressing the Quality Change Issue in the Consumer Price Index," *Brookings Papers on Economic Activity I*, 1997, PP. 305-366. Unlike the CPI, which compares prices over time, the index used here compares prices in a single period – expected, future prices given the innovation versus hypothetical, future prices assuming no innovation.

achieve the relatively superior performance of the innovations.

The cost index is constructed as the geometric mean of a Laspyres index—measuring consumer willingness to accept compensation to give up the gains from the innovation – and a Paasche index, measuring their willingness to pay to receive the gains from innovation. Both are measured relative to the baseline, and neither is theoretically superior to the other. The Tornqvist index is an equally weighted geometric average of the two.[31]

Austin and Macauley applied the method to evaluate two ATP-funded innovations in data storage technology. One innovation, undertaken by LOTS Technology, Inc., is an optical tape read/write technology representing a dramatic increase in data storage capacity. The second innovation, carried out by Imation Corporation, is to develop underlying technology for a linear scanning of magnetic tape that, at a fraction of the cost, can match or exceed the performance and capacity of a helical-scan system, a competing technology led by offshore competitors. Both technologies promise dramatic price/performance improvements compared to existing tape drives, but neither has yet come to fruition.

Using the method, the researchers estimated the expected net benefits to consumers from the optical tape technology to exceed $1 billion, and from the linear scanning technology, $2 billion, both taken over a five-year period. This analysis focuses on consumer benefits from early commercialization of the technologies, and ignores benefits accruing to the innovator or to other manufacturers via knowledge spillovers, as well as benefits from second-generation products.

The cost-index approach is a potentially useful tool for resource allocation by R&D managers in both the private and public sectors. One of the model's strengths is that it incorporates information heretofore difficult to obtain, and it varies all of the parameters simultaneously, making it possible to see the implications of changes to the parameter values and assumptions within a unified framework. The new model can be used to evaluate a single project's consumer benefits over time, and also to compare the future benefits of projects that may be competing for private-sector or government R&D funds.

A weakness is that uncertainties and omissions may cause actual outcomes to differ from the forecasts. Benefits from knowledge spillovers are not included in the model; only market spillovers are included.

Selected statistical/econometric studies not yet published

A number of the studies commissioned by the ATP that employ statistical and econometric methods are still outstanding at this time, without published reports. Several of these are summarized briefly to indicate further the ATP's use of these methods in its evaluation program.

- Professor Nicholas Vonortas of George Washington University conducted an empirical

[31] The theory of index numbers indicates that no single index satisfies all "desirable" properties or tests, e.g., tests related to scalability, transitivity, symmetry, and proportionality. The Tornqvist index satisfies many of the tests. For the mathematical formula the researchers use to construct a Tornqvist cost index, see the larger report from which this description is drawn, D. Austin and M. Macauley, *Estimating Future Consumer Benefits from ATP-Funded Innovation: The Case of Digital Data Storage*, NIST GCR 00-790, Gaithersburg, Maryland, 2000.

appraisal of the effectiveness of the ATP to foster joint research ventures. He performed statistical analysis of two databases – one from the U.S. Department of Justice's National Cooperative Research Act and one from the Advanced Technology Program – to compare ATP joint-venture formations with non-ATP joint venture to test the hypothesis that the collaborations ATP thinks it is fostering would have occurred anyway. He concluded that the hypothesis did not hold, and ATP was fostering the formation of joint ventures. (Unpublished ATP report)

- Professors Lee Branstetter of UC-Davis and Mariko Sakakibara of UCLA used econometric methods to assess the impact of ATP research joint ventures on R&D productivity of project participants. The study used patents granted as the measure of R&D output. They tested the model using Japanese patent data. They then used limited ATP patent data to investigate the impact of ATP-funded research joint ventures on firm patenting, and found a positive effect. (Unpublished ATP report)

- At the time of his recent death, Professor Zvi Griliches of Harvard University was working with Mauel Trajtenberg of Hebrew University in Israel, and Haim Regev of the Israeli Office of Statistics on an econometric study for the ATP. The researchers were developing an approach for analyzing the effects of government R&D subsidies on firm performance using empirical data from a similar Israeli program for which there is a much longer history and more data than for the ATP. The study was to draw on the available longitudinal data to demonstrate that the impact on firm performance and productivity of government R&D funding awarded to the firms could be successfully analyzed. The study was to examine whether government subsidized R&D differs from company privately financed R&D in its impact on firm performance. The study was also to analyze how various program mechanisms and instruments operate and how they contribute to program goals. A broad objective was to improve over previous studies of productivity of government funding which tended not to distinguish among types of government funding. (Unpublished ATP report)

Summary and Conclusions

Professor Irwin Feller, quoting Tolstoy at a recent workshop sponsored by the National Academies of Science, stated, "Doing good will not make you happy, but doing bad will surely make you unhappy." To this, he added the evaluator's corollary: "A good evaluation showing bad results will surely kill a program, but an evaluation that shows good results may not save a program."[32] Thus, even though the National Academies have concluded that ATP is "a federal program meeting uniquely challenging goals" and ATP's body of evaluative work can document this better than perhaps any research funding agency ever has, political opposition to ATP may prevail in the end. Nevertheless, the body of evaluative research conducted for ATP over the past decade has advanced the state-of-the-art in economic methods applied to evaluation and sets high standard for those who follow.

There are several reasons that likely explain why the ATP has developed a particularly strong evaluation program. One reason is that, unlike many other agencies of government, ATP's evaluation program was not a reactive response to the GPRA; rather it was out in front of the

[32] National Academies of Science, National Research Council, Board on Science, Technology, and Economic Policy, *Government-Industry Partnerships: The Advanced Technology Program; An Assessment of the Advanced Technology Program* (Washington, D.C.: National Academy Press, 2001),

GPRA. ATP established its evaluation program because its management believed it would be useful, and this predated both the GPRA and attacks on the program, which began in 1994. The evaluation program was largely motivated by the desire of those building and operating the program to make it work better by understanding as much as possible about the relationships among various program inputs, outputs, and outcomes. At the same time, realizing that program critics would subject all the ATP's evaluation results to intense scrutiny no doubt contributed to the program's increasing attention to rigor.

A second reason for the strength of ATP's evaluation program was that the office in charge of it paid close attention to following best practices of evaluation. In part, this was influenced by the NIST culture that puts strong emphasis on striving to be the best at what it does. One aspect of following best practices has been to maintain credibility for the evaluation program. By gaining the interest of leading economists and others in the field of technology evaluation, and demonstrating a willingness to fund sophisticated, state-of-the-art evaluation studies, the ATP has been able to attract the talent needed to build a substantive and credible program in evaluation. This effort included resisting occasional efforts by others to bring "PR" activities into the realm of evaluation. It has also included obtaining peer review of evaluation studies proposed and completed.

Evaluators have made strong use of several techniques in evaluating the ATP, particularly the survey method and the case-study method using benefit-cost and statistical analyses. It has found the descriptive richness of the case-study method valuable for generating interest in technology innovations among diverse audiences, and linking the cases to quantitative tools, such as statistical and benefit-cost analyses, to provide useful evidence of outcomes. It has made the use of counterfactuals and control groups standard practice in its evaluation studies. By establishing specific tests which ATP needs to meet for success, the evaluation program fostered awareness of success criteria within the program itself, and provided a basis for knowing when it had succeeded against theoretically correct goals.

Finally, the ATP has provided fertile ground for evaluation. Staff economists and academic and consulting economists working with them have seen the strong potential ATP offers as a laboratory for extending the state of knowledge of innovation and understanding better how to foster it for economic benefit through private-public partnership programs.

Chapter III. Bibliometrics

by Diana Hicks and Francis Narin

Introduction

Bibliometric data offer unique added value compared to other quantitative data on R&D such as funding, numbers of students etc. With all quantitative data one can produce counts, but bibliometric data also carries citations, and citations have a variety of uses. For evaluation, the count of citations received by the work being evaluated is paramount. A citation suggests linkage, and there are three types of linkage data useful for evaluation: paper-paper, patent-patent and patent-paper. Because the linkage information is unique to bibliometric data, methodological issues in the treatment of citations are not widely appreciated and in the absence of understanding, criticism is sometimes offered. This chapter explains the background and use of all three types of citation data in the evaluative context, setting out the development of the methods, briefly describing validation studies, discussing strengths and weakness and elements of best practice for evaluation purposes.

History and validation

Origins of Science Citation Analysis

The origins of the idea that citations would provide a method of gauging the impact of a scientist's work are apparent in the following quotes from Garfield's 1955 paper in *Science*.

> "Since 1873 the legal profession has been provided with an invaluable research tool known as *Shepard's Citations*, published by Shepard's Citations, Inc., Colorado Springs, Colo. (2). . . .Some time ago I became concerned with the problem of developing a citation code for science. This was necessary for the efficient manipulation by mechanical devices of entries to scientific indexes. . . This would clearly be particularly useful in historical research, when one is trying to evaluate the significance of a particular work and its impact on the literature and thinking of the period. Such an "impact factor" may be much more indicative than an absolute count of the number of a scientist's publications.
>
> (Garfield, 1955, p. 108-109)

In the early 1960's Garfield created the *Science Citation Index* which has grown to cover more than 5,000 scientific journals, more than half a million papers per year, and more than 5 million citations annually.

Although Garfield and his colleagues were well aware of the potential use of citation data in measuring the impact of individual papers, the widespread acceptance of science citation data in evaluation is associated with the creation by the National Science Foundation of the first *Science Indicators 1972* report. Narin and his colleagues at CHI Research (CHI, then called Computer Horizons, Inc.) took the *Science Citation Index* data and created national and international

scientific performance indicators. They used counts of publications, and most importantly, counts of how frequently those publications were cited, to create the first bibliometric indicators of national scientific performance, which were used in that report. That effort has continued biennially as the now standard indicators are produced and new indicators are developed.
In the late 1970s and early 1980s CHI developed the citation analysis techniques for research evaluation. Much of this development was done for the National Institutes of Health in studies summarized in Chapter 4. This work was widely reported at the time, for example in reviews of evaluations.

In succeeding years, the techniques developed for science citation analysis were further developed for use in patent citation analysis. The precipitating event for this transfer of technique and knowledge from the scientific realm to the patent realm was the release, in the early 1980's, of the U.S. Patent Office bibliographic tapes which contained not only the titles and abstracts of the U.S. patents issued, but also all of the ‘references cited’ on the front pages of those patents. With this database, CHI constructed an analytical patent citation indicators database. The analysis techniques used with this database have been greatly expanded in the ensuing 20 years, and are becoming more and more precise. The initial uses of patent citation indicators were analogous to the initial *Science Indicators*, measures of national and international performance. The patent indicators were then taken forward from this policy level to a much more precise strategic level, dealing with measures of the productivity, impact, and importance of technology at the company level. These were then further refined to the tactical level, for use in tracing the development of individual technologies, and for technology analysis. This work has been drawn upon by government agencies interested in their output of patented technology.

The most recent development of citation analysis is in the area of patent-paper or linkage (that is science-technology linkage) analysis. In the early 1980s, CHI noted the growth in references from patents to non-patent material, including scientific papers and sought to extend its citation analysis techniques into this new area. Since then the amount of referencing from patents to papers has grown dramatically making available a new indicator of research outcomes – namely the number of times papers are referenced in patents. In the late 1990s, CHI produced a widely reported analysis which found that 75% of the papers cited by patents were produced by public sector institutions. CHI has also conducted many linkage studies for government agencies and these are summarized in Chapter 4.

The development of bibliometrics from paper-paper to patent-patent and finally to patent-paper has taken several decades. Along the way, methodological innovations have been accompanied by studies that examine the correlates of high citation and the meaning that can be given to the indicators. These studies will be discussed in the next section.

Validation of Science Citation Analysis

Most of the studies examining the validity of science citation analysis have been done by means of correlation, studying the correlation between high citation, and many other available but still relatively informal measures of scientific accomplishment and quality. There are, of course, many thousands of papers dealing with science citation analysis, hundreds of which deal directly with these validation concepts. Approximately 24 of the early papers were classified, reviewed, and discussed in Chapter V “Correlations with Non-Literature Measures” of CHI’s Monograph *Evaluative Bibliometrics*, which reported for the National Science Foundation on the early use of bibliometric techniques in research evaluation. (Narin, 1976)

The early validation techniques covered the full range of studies being used in research evaluation. They covered correlation between publication and citation measures of national, institutional, research group and individual performance. At the national level, for example, some very early work by Derek Price showed that nations publish roughly in proportion to their Gross Domestic Product (GDP): that is in proportion to their economic size, not to their population or land area or anything else. Much later CHI has shown that this also carries over into technology, and that other nations' inventors patent in the U.S. patent system in general proportion to their national economic size as measured by GDP (Narin, 1991)

At the institutional level citation techniques have been applied extensively to the ranking of university departments, which has been done systematically in the United States in a series of reports in which relatively large numbers of senior academics voted on the relative ranking of major departments. In a paper published in 1978 and reprinted in 1980, CHI showed that not only do these peer rankings of universities correlate well with publication rankings, but that the correlation always increased substantially when citation data is included. In other words, the ranking of a university based on a combination of number of papers and how frequently they are cited is much more highly correlated with peer rankings than one based on publication frequency alone (Anderson, Narin, McAllister, 1978).

Citation techniques are also used for measuring the importance of individual research accomplishments. These techniques have been used in studies of the correlation between citation records and faculty positions in university departments and prestigious awards, and with other measures of prominence and acceptance within the scientific community. In many of these cases it has been shown that there is a strong correlation, generally in the range of 0.5 to 0.7, between independent measures of productivity and quality in science, and measures developed from publication and especially citation techniques.

One of the most fascinating and telling demonstrations of the importance of very high citation is a series of papers that have been published related to the bibliometric characteristics of Nobel laureates in science. In a paper discussing the quality of research and Nobel prizes Inhaber notes: "The quality of the work of Nobel laureates in Physics, as measured by citations, is an order of magnitude higher than that of other scientists" (Inhaber and Prednowek, 1976, Page 34).

Basics of Patent Citation Analysis

When a U.S. patent is granted it typically contains seven or eight "references cited-U.S. patents" on its front page. These references link the just issued patent to the earlier cited prior art, and limit the claims of the just issued patent. They point out where essential and related art already exists and therefore limit (are cited against) the claims of the citing patent.
In one important respect patent citations differ from citations in a scientific paper. Front-page patent references are either put there by the examiner, or suggested by the applicant and his attorney and approved by the patent examiner, who is an expert in the art which he is examining. The net result of this is that a patent citation is undoubtedly stronger and more likely to be relevant to the subject area of the patent than a literature citation.

When this referencing pattern is turned around, and all of the subsequent citations to a given patent are tabulated, one obtains the fundamental information used in patent citation analysis, namely how often a given patent is cited in later patents. These distributions tend to be very skewed: there are large numbers of patents that are cited only a few times, and only a very small number of patents cited more than once. Half the patents are cited 5 or fewer times, and only one percent of the patents are cited 48 times.

Validation Studies for Patent Citation Analysis

In this section we will discuss various studies, which relate higher than average citation frequency for patents to other indicators of patent importance, such as the opinions of peer scientists, awards for technical excellence, and the legal status of "pioneering patent" awarded by U.S. courts. The first paper that looked at patent citations as a way of finding important patents was a very early study done by Reisner at IBM, who experimented with the use of citation analysis to find key patents (Reisner, 1965). By tracing the references from one patent to another, Reisner found 47 of 60 key patents she was looking for. Complete citation data covering all U.S. patents issued first became available in 1975. In the following year, in the sixth Technology Assessment and Forecast report, the Patent & Trademark Office tabulated the patents which were most highly cited, and suggested that "the number of times a patent document is cited may be a measure of its technological significance" (PTO, 6th Report, 1976). In the mid-1970s, Ellis, Hepburn and Openheim in the United Kingdom experimented with patent citation networks, tracing the references from sets of patents to see if they could identify key discoveries and key turning points in a group of technologies (Ellis et al 1978). The analysis was relatively successful, in that for the cases in which there was a very key starting event, that key discovery was often apparent from the patent clusters. On the other hand, in areas where a technology developed over a relatively long period of time, that was reflected in a less focused pattern without a key cluster.

The first relatively formal study of patent citation analysis was carried out by CHI Research under the sponsorship of the National Science Foundation (Carpenter et al, 1981). At the time the study was proposed, in the late 1970's, the *Science Indicators Unit* at the National Science Foundation was considering whether to add technology indicators based on patent citations to the stable of science literature indicators, which were then being used in the *Science Indicators* reports. NSF commissioned CHI to do a study to see whether patents associated with important discoveries were more highly cited than average patents. The set of important patents was obtained by attempting to determine the key patent underlying a product which received the IR-100 award established by the journal *Industrial Research and Development*. This award

> "honors the 100 most significant new technical products -- and the innovators responsible for them -- developed during the year. From thousands of entries, the distinguished Editorial Advisory Board of *Industrial Research* selects the 100 products that are most important, unique, and useful. Extensive local and national press and television coverage of the winning entries and awards presentations has made the IR-100 award the most coveted achievement in the applied research and development field" (Industrial Research & Development, 13, p3, December 1980).

Patents related to the 1969 and 1970 awards were used in order to ensure that there was sufficient time for the patents to be cited to their full potential. A set of 100 important patents and a set of 102 control patents were selected. The study found that the IR-100 patents were cited on average twice as often as the controls (4.9 cites/patent vs. 2.0 cites per patent), and among the IR-100 patents there were 17 cited more than 10 times compared to 4 such highly cited patents among the controls. Clearly, the IR-100 patents are much more highly cited, and much more likely to be very highly cited; this difference is due to the presence of highly cited patents in the IR-100 set.

Subsequent to this study, patent citation indicators were added to the *Science Indicators* reports (by then called *Science and Engineering Indicators*), and their use has expanded since.
A somewhat different approach, with the same results, was taken in a study by Trajtenberg, with the marvelous title "A Penny for Your Quotes" (Trajtenberg, 1990). Trajtenberg analyzed patent citation patterns associated with advances in CAT scanners, and showed a close association between citation-based patent indices and independent measures of the social value of innovations for computed tomography scanners. Of particular significance is high finding that "the weighting scheme appears to be non-linear (increasing) in the number of citations, implying that the informational content of citations rises at the margin" (p. 172). This directly supports the idea that highly cited patents are of particular technical importance.

CHI Research in cooperation with Eastman Kodak Laboratories investigated the possibility of using patent citation data in analysis of their technology and of some of their competitor's technology. Kodak wanted to independently validate whether, within an industrial laboratory, high patent citation was associated with knowledgeable peer assessment of the importance of the patents. A carefully designed study was carried out in which sets of 16 Kodak patents, from Kodak's core area of Silver Halide Technology, were given to 20 senior lab staff for evaluation. The sets were overlapping so that every patent was in at least three or four different sets, so that the rankings of the patents could be cross tabulated. The Kodak evaluators were senior intellectual property staff, senior lab management, and senior lab scientists. In the case of scientists, the patents they were given were screened to make sure that they did not rank their own patents. Each person was asked to rank the patents from the patent that had the most technological importance to the patent that had the least. The result was that whether a patent is cited one, two, or three times does not seem to make much difference in the peer ranking. However, patents that are cited more than five times, that is, relatively highly cited patents, were ranked far more highly by the Kodak staff. Of the 15 respondents represented in the study, eight gave the most highly cited patents the highest average rating. Using a binomial model, the probability of this is 0.0002.

The most recent evidences for the importance of highly cited patents comes from within the Patent Office itself, in the form of strong associations between citation frequency and Patent Office recognition, and in the extremely high citation to pioneering patents. CHI has looked at the citation frequency of three different categories of patents: patents listed in the National Inventor's Hall of Fame, patents of Historical Significance in a list prepared by the U.S. Department of Commerce for the U.S. bicentennial, and patents that had been adjudged as pioneering patents by the Federal District Court.

> "The National Inventors Hall of Fame is dedicated to the individuals who conceived the great technological advances which this nation fosters through its patent system.
> Inventors are selected for The Hall of Fame by the Selection Committee of the National Inventors Hall of Fame Foundation. The Selection Committee is composed of representatives from national scientific and technical organizations. Each year the members of this Committee vote to select the most qualified inventors from those who are nominated for induction.
> In its voting, the Selection Committee considers whether the invention of the nominee is covered by a United States Patent, the contribution of the invention to the nation's welfare, and the extent to which it promotes the progress of science

and useful arts."

(National Inventor Hall of Fame, 1993)

The study found that most of the Hall of Fame, Historically Significant and Pioneering patents were cited well above the average, with only a single Historical Patent below the average in citation frequency. The Pioneer Patents were cited five times more than an average patent for a given year.

Financial And Economic Correlates Of High Patent Citation

Another set of illustrations of the association between patent citation rates and the quality of technology may be inferred from the associations being uncovered between economic and financial data, and patent citation counts.

By the early 1980's, it was possible to obtain reasonably large scale patent data, and Griliches and his colleagues at Harvard, and the National Bureau of Economic Research (NBER) began a long series of quantitative studies looking at the economic importance of patents. In a 1981 paper Griliches found a significant relationship between the market value of the firms, and its "intangible" capital, provided by past R&D expenditures and the number of patents (Griliches, 1981).

In 1987, Narin and his colleagues studied a group of 18 U.S. pharmaceutical companies, and showed that the numbers of patents they obtained and especially whether the companies had highly cited patents, were both correlated with peer opinions of the companies, and with increases in pharmaceutical company sales and profits (Narin, Noma, & Perry, 1987). That study showed quite clearly that highly cited patents tended to occur around economically important inventions such as Tagamet for SmithKline, and that these important technological events lead, in that industry, to increases in company sales and profits.

In 1992 and 1993 Business Week published two Patent Scoreboards using CHI Research data to rank major companies across 10 different industries (Buderi et al 1992; Coy & Carey, 1993). The Patent Scoreboards were two of the first times when these ideas were introduced directly to the business community, so that analysts could look at the relationship between the business performance of companies and their technological strengths.
The economists associated with the NBER are now using patent citation techniques in a wide variety of ways, studying spillovers of research from company-to-company, and university-to-company, studying the characteristics of successful companies, and in general, demonstrating the acceptance of the notion that patent citation is equivalent, in the statistical sense, to high impact technology. A paper by Jaffe, Trajtenberg and Henderson (1993) provides a linkage into this literature.

A recent NBER report "Market Value and Patent Citations: A First Look' finds that "citation-weighted patent stocks are more highly correlated with market value than patent stocks themselves and that this fact is due mainly to the high valuation placed on firms that hold very highly cited patents" (Hall, Jaffe and Trajtenberg, 1998)

A recent paper by F.M. Scherer of Harvard, and colleagues in Europe and at CHI, looked at a sample of U.S. and German patented inventions on which profitability information – that is the private value of the patents – was obtained (Harhoff, Narin, Scherer, & Vogel, 1999). They considered only patents for which all the fees had been paid to keep the patents in force in

Germany for the full 18 years of the patents, and then queried the owners of those patents as to the asset value of the patent – essentially asking, what is the smallest amount they would have been willing to sell this patent to an independent third party for in 1980? In the German patent system the two patents in the highest value category were much more highly cited than the others. In the US patent system the patent citation frequency of the patents with an estimated value of $20 million or above were substantially more highly cited than the patents with lesser estimated values.

Deng, Lev and Narin have looked at the relationship between patent citation indicators, and various financial indicators, including R&D budgets, and stock market performance (Deng, Lev, and Narin, 1999). In particular, they found companies whose patents had above average citation (Current Impact) indices and science linkage indicators (SL's) tended to have significantly higher market-to-book ratios, and stock market returns, both contemporaneously and for a number of years into the future.

Thus it is not surprising that CHI has developed methods of using patent indicators to select high performing stock portfolios. CHI has patented its methods of using patent citation techniques to identify companies whose stock market value is likely to increase in the next few years.

Patent-paper bibliometrics

For more than two decades, CHI Research has been analyzing the references made by U.S. patents to the scientific literature, investigating their use as indicators of the linkage between science and technology (Carpenter et al, 1980). The first study of references in patents to scientific papers sought to extend methods of bibliometric analysis developed for the scientific literature. The study investigated two fast growing areas of patenting closely linked with scientific research - prostaglandins and gas lasers. The purpose of the study was to discover whether patents in areas where technological development was known to be closely linked to research would show evidence of those close links in high levels of referencing to scientific literature. The results showed that the patents in these science-linked areas did indeed reference substantial amounts of science (Carpenter & Narin, 1978).[33] In a subsequent study, 19 high level R&D managers were asked to rank 24 technologies by their science dependence, and their rankings exhibited a high degree of agreement with the rate of referencing to papers (Carpenter and Narin, 1983). CHI pursued this promising technique, building the databases needed to move from case study to systemic level analysis. Therefore, we know that the referencing from patents to papers is stronger and more broadly based today than in 1978. The most science-linked area is biotechnology where patents reference an average of 20 scientific articles.[34] Concomitantly, in areas known not to be particularly research intensive the patents do not reference much scientific literature - industrial machinery and tools or textiles and apparel for example.[35] CHI has argued that this type of citation data is unrivaled in enabling

[33] Both examiner and applicant references were examined. In prostaglandin patents SCI journal articles were referenced at a rate of 8.1 per patent = 0.5 (examiner) + 7.6 (applicant). In gas lasers, the rate was 1.9 per patent = 1.1 (examiner) + 0.8 (applicant).

[34] U.S. invented patents assigned to companies and issued in 1999. Examiner (front-page) citations only.

[35] Examiner (front-page) citations to scientific literature on U.S. invented patents assigned to companies and issued in 1999: industrial machinery and tools - 0.22, textiles and apparel - 0.37.

comparative quantitative analysis of science and technology linkage (Narin and Olivastro, 1992[36]).

In 1997, Narin and colleagues published an influential study of citation from U.S. patents to U.S. papers. The paper analyzed the 430,226 non-patent references (NPR's) which were listed as 'other references cited' on the front pages of the 397,660 U.S. patents issued in 1987-88 and 1993-94. Of the 430,000 NPR's, approximately 242,000 were judged to be science references - that is, citations to scientific journal papers, scientific meetings and other scientific publications. Of these 242,000, about 175,000 were references to papers indexed in the *Science Citation Index*. References that could be matched to items indexed in the SCI were further winnowed to include only papers with U.S. authors, published in the 11 years prior to the issue year of the citing patent. 45,000 cited papers were left, and these were looked up in libraries to obtain sources of support acknowledged by the authors.

The study concluded that public science plays an essential role in supporting U.S. industry, across all the science-linked areas of industry, amongst companies large and small, and is a fundamental pillar of the advance of U.S. technology. Furthermore, the data showed that the science that contributes to high technology is mainstream, it is quite basic, quite recent, published in highly influential journals, authored at major universities and laboratories, and supported by NSF, NIH and the Departments of Defense and Energy, and by other public and charitable institutions.

The study was reported in the New York Times (see appendix for reprint). It was also said by insiders to have been influential in budget deliberations in the White House and Congress. More recently, a study was undertaken for the Australian government which has also had influence on government decision making in science and technology. This study is described in Chapter 4.

Strengths and Weaknesses

Paper-Paper bibliometrics

The complexities of bibliometric analysis are many and varied. Some difficulties are technical in origin and can be resolved by careful data compilation; others originate in scientists' citation practices but can be dealt with through the appropriate design of the research and appropriate interpretation.

Attempts to measure the scientific literature are now possible thanks to the Institute of Scientific Information (ISI), Philadelphia, USA, which compiles the details of papers in thousands of journals and makes them available in the *Science Citation Index* (SCI). Technical difficulties in using the SCI fall into three categories: coverage, consistency and first-author indexing. Although ISI scans a great many journals, a vast number are omitted and books and their reference lists are mostly excluded from the database. ISI includes a number of 'core' journals, which it argues constitute the most significant part of the world's literature. This produces a bias against non-English language literature – which means that from the Japanese perspective, SCI citation and publication counts can provide fair assessments only for internationally oriented work.

[36] This paper contains a detailed review of CHI's linkage analyses up till that time.

Problems of consistency concern errors in data recording. Mistakes might be made by the authors of the publications themselves, by the publishing journal or by ISI incorrectly inputting citations and publications to the database. If sufficient attention is given to these problems, for example if a high quality algorithm is used to match citations with cited papers, bibliometric analysis can be very accurate.

The third problem arises because citations list the first author only. Looking up ‘D Smith’ will not locate all of Smith’s publications and will include cites to work of other authors with the same last name and first initial. The most satisfactory solution is to work from complete bibliographies and to match cites to full paper references.

Other problems originate in scientists’ publication and citation practices, but are nonetheless amenable to technical solutions. The first is bias in SCI against applied research, particularly areas such as manufacturing engineering that are not closely related to scientific disciplines. This is partly a result of the SCI’s incomplete coverage: ‘gray’ literature and patents are excluded. The problem runs deeper, however, for it is simply impossible to use published output to assess or understand research where an immediate goal is to develop devices or improve technique. With some exceptions, paper-paper bibliometric analysis can be best used to assess contributions to the advance of knowledge in areas where articles in international scientific journals are important.

Bibliometrics is further complicated by the variations in publication and citation practices between different fields. Citation rates vary widely between fields as illustrated in Figure III-1, which plots the average number of citations per paper across a variety of subfields. In general, biomedical fields such as virology tend to be highly cited while engineering areas like aerospace technology or mathematical areas like probability and statistics tend to have lower average citation rates. To be meaningful, studies must choose one of three strategies. First, studies can compare similar groups within the same subfields: in other words, they must compare like with like. Alternatively, publications can be ’weighted’ to reflect the prestige of journals in which a particular group publishes - referred to as "influence weights" below. Normalization is another alternative, in which citations earned are compared with the average number of citations to articles in these journals.

The rate at which scientists cite themselves is the final problem in this category, and individual cases must be inspected to ensure that variations in self-citation do not affect conclusions drawn from the data.

A third category of problems with bibliometric analysis calls for an understanding of why scientists publish in the way that they do. Critics of bibliometrics refer to ‘overciting’ and underciting’, and speculate widely about ‘illegitimate’ referencing behaviors. When papers receive more citations than they ‘should’ in relation to their intrinsic scientific worth, we might say that they are ‘overcited’. Some work is cited by other papers pointing out errors or refuting arguments, so citation counts can measure something other than the quality of a piece of work or its contribution to science. A similar problem is that papers about methods or techniques are overcited in relation to what some consider to be their scientific merit. The most cited paper of all time, by Lowry, is a methods paper. Finally, if an author published several short or repetitive papers which contain essentially the same idea, another author may cite them all, thus creating the appearance of inflated intellectual influence.

The opposite effect, namely that papers are cited less than they ‘should be’, also has several sources. The best known types are commonly called ‘neglect’ and obliteration’, an example of the first being Mendel’s work, ignored until many years after its publication, and an example of the second being Einstein’s $E = mc^2$, so universally accepted that the original paper is no longer explicitly cited. The injustice of obliteration is illustrated by a study of mathematics which found that older papers that made major contributions are now cited less than papers of a similar age that made minor contributions.

Further reasons why the references at the end of a paper might not record all the intellectual influences on the work include: basic assumptions and background knowledge may not be referenced, more recent influences may not be referenced, parsimonious citation of review papers or schools eliminates citations to the individual papers in which ideas originated, and informal influences are not cited. There are also omissions resulting from ignorance, memory failure, lack of awareness, or carelessness.

A final criticism is that citations may be unrelated to the scientific content of the cited paper. Visibility is said to enhance the likelihood of citation: the ‘elder statesmen’ of science find it easier to publish and are ritually cited, in much the same way that classics are cited to show competence. There are a host of other reasons for self-serving citation practices, contributing to a general concern that referencing is not a wholly ‘rational’ process and that this makes citation counts unreliable as indicators of scientific quality or importance.
With such an extensive list of potential sources of irregularity at the micro-level of citation practice, one wonders how it is that any patterns and trends emerge from bibliometric analyses. But the fact remains that they do, and that often they make sense to experienced interpreters. As reported earlier, the relationship between bibliometric indicators and traditional indicators of scientific excellence has been well studied. Correlations have been found with such measures as peer evaluations, ratings or esteem, and departmental rankings. But this begs the question why this should be so, and moreover, what should be done if a bibliometric analysis contradicts scientists’ perceptions? There are a number of points to consider.

Firstly, most citation and publication studies adopt relative not absolute measures of scientific performance. For example, the specific characteristics of scientists’ behavior that are common to all the laboratories in a particular field should not necessarily affect *comparisons* between these laboratories. The higher the level of disaggregation, of course, the greater the possibility that local factors will interfere. Secondly, citations and publications cannot be interpreted as straightforward indicators of scientific merit. There are four terms that are usually confused when describing the characteristics of scientific work that bibliometric indicators measure: quality, importance, impact, and citation (or publication) rates. ‘Quality’ describes how well the research has been done; it is a matter of judgment, and will therefore be evaluated differently by different people at different times. ‘Importance’ is the potential influence of a paper on a specialty: because of the imperfections in scientific communication, the *actual* influence diverges from this potential and is called ‘impact’. Each of these terms encompasses more contingent, external factors than the previous, including, for example, how well written the paper is, the eminence of its authors, their reading or referencing habits, and the size and dynamics of the field. Since further social factors intervene between impact and citation or publication rates, all one can measure are the latter, which are only partial measures of scientific impact. (See Martin & Irvine, 1983 for a full treatment of this issue.)

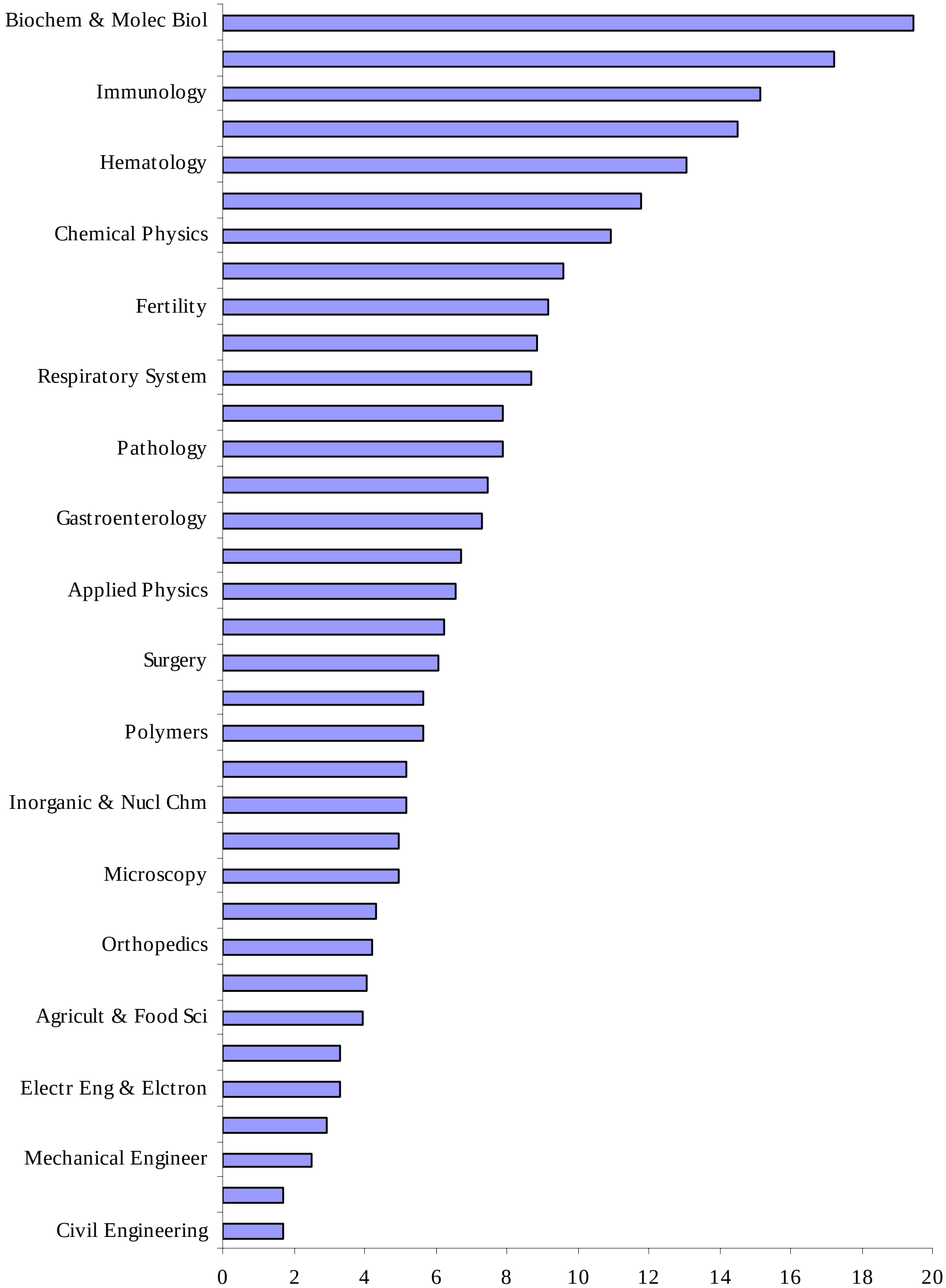


Figure III-1. Citations per paper vary by subfield

Average number of citations earned by 1989 papers in the 5 years following publication using a 1997 fixed journal set

Applying this scheme to bibliometrics resolves many of the problems outlined above. Citations and publications are indicators of recent activity in scientific communication, and it is therefore not surprising that they often coincide with scientists' own perceptions of the shape and distribution of their field. These partial indicators, however, are evidence of quality or importance only when they have been demonstrated to be so. They should therefore be complementary to other methods of evaluation, not substitutes for them.

Experience in the UK tends to support this contention. Agencies using bibliometrics found that the main benefit of such indicators is to focus attention rapidly on areas of low or high productivity, and to pose questions that otherwise might not be obvious. Quantitative data can force 'difficult' funding decisions, preventing the peer review system from becoming too soft. In contrast, where peer review conflicts with the evidence, experts can be required to justify their position, thereby making the decision-making process more transparent.

In general, these research funding agencies found that performance indicators should not be employed mechanistically, e.g. as variables in fixed-funding formulae. Rather they were most successful when individual scientists and institutions were involved in their construction and interpretation. Not only did a consultative approach produce an atmosphere of openness, reducing the hostility that inevitably follows any exclusively 'top-down' approach, but it also enabled researchers to benefit from the review process themselves.

To summarize this discussion, two tables are used. In Table 3-1, commonly mooted problems with citation analysis are listed along with methodological strategies that minimize or eliminate the problems. In Table 3-2, the strengths of bibliometric analysis are reported along with an explanation of why each element is important in the evaluative context.

Patent-Patent Bibliometrics

Methodological problems in the use of patent citation analysis to evaluate the outcomes of public research will not be discussed in the same detail. Suffice it to say that similar questions arise and in large measure similar answers apply. Specifically, all three issues listed under "publication practice" in Table 3-1 apply to patent analysis and the solutions are analogous. Some problems of "referencing and citation behavior" apply, namely numbers 1, 4, 5 and 7. The solutions are analogous, except for number 7 - time lag, because with patents there is no proxy measure of quality, such as journal influence, that is immediately available.

Table III-1. Main problems with bibliometric indicators and how their effects may be minimized an update of Martin & Irvine's classic table (*Research Policy*, v. 12, 1983, p. 76)

• **Category** Purported problems	**Is it really a problem?** **How is it minimized or eliminated in properly conducted studies?**
• **Publication practices**	
1) propensity to publish varies by type of institution, field and country	Matched groups can be compared or normalizations can be devised at the institutional, field or country level. Do not rely solely on bibliometrics to evaluate institutions for whom publishing is secondary.
2) manipulation	Bibliometric indicators are more difficult to manipulate than some other evaluative techniques. Some forms of manipulation are counter-productive when citation rates are also analyzed - for example, producing more shorter papers will more than likely reduce average citations per paper and the number of highly cited papers. Other supposed manipulations are simply rational choices whose impact will be felt in any evaluation - for example, avoidance of low prestige, low impact areas where publication is difficult.
3) does not indicate quality	Use citation-based measures to indicate impact
• **Referencing and citation behavior**	
1) varies by subfield	Raw counts can be used to compare matched groups, for example within one specialty. Most evaluations require normalized citation measures that permit comparison between specialties.
2) varies by type of contribution - review, method, experiment, theory	Matched groups can be compared. Regard citations as indicators of impact. Also note that not every review or method is highly cited, nor is every theory uncited.
3) critical citations	Regard citations as indicators of impact rather than quality or importance.
4) self-citation	Only highly productive scientists can achieve high rates of self-citation (since one needs many papers to generate many self-cites). Therefore, self-citations tend not to change the results of evaluations. If resources are available, self-citations can be removed from consideration.
5) citation rate varies over the lifetime of a paper	Citations to papers peak at 3 to 4 years. Counting cites to each paper in a fixed window, such as the first four years after publication ensures each paper is given the same chance to acquire citations. The impact of "sleeper" research, whose impact is manifest much later will not be picked up by bibliometrics, or any near-term evaluation.
6) inflexible with regard to special circumstances	Authors or institutions may be in special situations that depress their publishing or citation counts relative to others doing equivalent work. Therefore, low scores should always be taken as a flag that *something* is unusual, and knowledgeable peers can be consulted to identify the reasons.
7) time lag	A measure of journal quality, such as journal influence, can substitute for citation counts if immediate evaluation is a priority.

• **Technical limitations of the *Science Citation Index***	
1) spelling and clerical errors	A high quality algorithm must be used to match reference strings with the papers referenced
2) citing 1st named author	Studies must begin by constructing a comprehensive bibliography and finding citations to the papers in the bibliography.
3) covers only journals	Absence of books may restrict application of bibliometrics to the social sciences (though this may be changing, see Hicks, *Scientometrics*, 1999). Comprehensive coverage of conferences would be needed to apply bibliometrics to software development.
4) journal coverage incomplete	Not a problem in many areas, particularly in basic research. Particularly problematic for technical areas not closely associated with science - such as production and manufacturing.
5) coverage varies over time	In situations where this becomes important, a fixed subset of journals can be used.

Table III-2. Main strengths of bibliometrics as used for evaluation

Strength of bibliometric technique	Importance
• **Methodological**	
1) quantitative	Not easy to manipulate. Will highlight potential areas of weakness for further investigation, not "soft". Precision and consistency.
2) can be normalized	The impact of those on the periphery, which might not be visible from the center, can often be seen with normalized indicators.
3) direct	Clear, concise, mostly simple counts. Few to no assumptions hidden in construction, any assumptions concern the meaning of citation counts and are known to all parties in an evaluation.
4) based on publication and citation	Because publication is central to scientific work and communication, measures derived from it are more difficult to influence than information constructed solely for an evaluation.
5) time lag	Since citations peak two to three years after publication, citation information may lag the award of grants by only five to six years, much less than the lag for technological application.
6) scales well	Large numbers of projects, institutions etc. can be examined without costs escalating out of control. The same indicators can be used at many levels of aggregation - from departments to countries.
7) studies assembled by those with no stake in the outcome	Greater objectivity, less influenced by partisan agendas and old boy network
8) unobtrusive	Places no demands of time or resources on those being evaluated. Thus when the full costs of evaluation are compared, bibliometrics is relatively economical.
• **Technical advantages of the *Science Citation Index***	
1) all addresses are indexed	Collaboration has increased so much that today assigning papers to the single institution listed in other databases is as inadequate as assigning a paper to its first author.
2) covers all scientific fields	Normalized indicators can be devised to enable comparison among fields on a broadly comparable dataset
3) covers top journals	The scientific literature is too vast for anyone to cover comprehensively. Choosing to cover the most highly cited 5,000 to 7,000 journals is a sensible way of choosing what to cover.
4) includes citations	Permits citation analysis.

Patent-Paper Bibliometrics

The power of patent-paper bibliometrics originates in its orientation to outcomes rather than outputs. One outcome from research is impact on commercialized innovations. Governments place great weight on this sort of impact, therefore being able to demonstrate it is valuable. However, it is difficult to demonstrate because it is famously diffuse and hidden by the passage of time; and therefore is expensive to uncover.

The increasing frequency with which U.S. patents reference papers has made visible some of the links between research and technology. Because a patent represents an identifiable intermediate step between a new invention and the later, downstream development of commercial applications, citations from these patents to the scientific literature provide quantitative and objective linkage between technological innovation and basic scientific research. The measure has been influential in several applications, however its interpretation is not without controversy in the academic community. In patent-paper bibliometrics it is important to understand: 1) the meaning of a citation on a U.S. patent; 2) the incentives in the U.S. patent system to disclose information, and 3) possible explanations for growth in patent-paper references.

The references to scientific papers analyzed by bibliometricians are found on the front page of U.S. patents among the "other references cited". These references, as well as references to prior patents, are required by patent law which stipulates that to be awarded a U.S. patent, an invention must satisfy three criteria: it must be useful, novel and not obvious to someone skilled in the art. References on the front page of the patent arise from the novelty requirement, since it is in the references that the patent applicants, their attorney and the examiner must identify all of the important prior art known to them upon which the patent improves. Front-page references are chosen and/or screened by the patent examiner, who is "not called upon to cite all references that are available, but only the best" (Patent and Trademark Office, 1995).

Using the linkage indicator with confidence requires an understanding of how the patent-to-paper reference it is based upon comes into existence. The requirement to document novelty explains part of patent referencing as explained above, but other circumstances also shape referencing behavior. The influence of these factors is particularly visible when U.S. and European patents are compared. European patents do not reference as much scientific literature, and have not greatly increased their referencing (Narin and Olivastro, 1998). Note that procedures for adding references to patents differ considerably between the two systems. In the US, the applicant supplies a list of prior art references with their application, whereas in Europe the applicant is not required to do this and all prior art searching is done by patent office staff. Thus, front-page reference lists on U.S. patents benefit from contributions from applicants aware of the full technological context whereas in Europe, front-page reference lists are constructed by patent office staff naturally focused on the patent system.

In the US, these procedural differences are reinforced by strong incentives. The U.S. rule of disclosure requires the applicant to disclose all known prior art references, and failure to do so is considered fraud on the patent office and can provide grounds to disallow the patent. Incentives to reference are strengthened in the U.S. in comparison to Europe by the heightened risk of patent litigation. The examiner is presumed to have examined a referenced document and to have decided that the invention was novel in relation to what was reported in the document. Therefore, a document referenced on a patent is much more difficult to use in court as evidence that the patent is invalid because the invention was not novel.

The procedures by which references are appended to patents combined with strong incentives to disclose prior art serve to make visible the patent-paper linkages in the U.S. system. Pursuing this further, we could ask whether examiners and applicants are pushed to place copious irrelevant material in patent reference lists. This seems unlikely because the strong incentives operate only concerning relevant prior art. Investing time and money in finding and appending references to irrelevant prior art affords no additional legal protection. Analysts who desire to characterize the relationship between science and technology are fortunate that the U.S. patent system provides strong incentives to document relevant non-patent prior art because among these "other references" on a patent can be found the references to scientific literature.

The patent office procedures and legal requirements surrounding patent referencing provide a sound foundation for interpreting the indicator as revealing science relevant to patented technology. Additional support for this interpretation lies in the pattern in which highly science linked technologies exhibit high levels of referencing to science, confirmed by experts. Note that interpretation of the indicator does not rest on a particular model of the social process through which references are appended to patents. For example, a fairly mechanical model of referencing behavior might be: inventor reads paper, inventor gets idea for invention, inventor files patent application referencing paper. That this model is quite unlikely to hold is suggested by Meyer's examination of 10 Norwegian and German patents (Meyer, 2000). That the mechanical model is unpromising is also suggested by studies of scientific referencing that could not find empirical support for similar rational mechanisms proposed to explain scientific referencing. Rational models of referencing lacked support in the scientific world even though they would seem to hold more promise there. After all, authors append the references to their papers in contrast to U.S. patent applicants who merely offer the examiner suggestions for front-page references.

Within the legal constraints, the social processes through which the close links between research and technology find expression in patterns of patent to paper citation are undoubtedly multiple and complex. In science intensive technologies, inventors may be more likely to have authored papers and thereby to have gained not only publications to cite but also a familiarity with the research literature which might enhance literature referencing on their patents. Examiners may be more likely to have Ph.D. qualifications because that is what is required to examine the science intensive patents. Ethnographic study of the construction of patents would be needed to tease out all of the contributing factors. However, the interpretation that the science intensity of a technology is reflected in the rate at which its patents reference scientific literature is not grounded on the operation of specific social processes - whether rational or not.

Interpreting the patent-to-paper citation indicator is complicated because complex social processes and patent office rules underlie the writing of patents. Interpreting growth in patent-to-paper referencing is even more complicated because it is possible that changes in the social processes and patent office rules move the indicator rather than changes in the relationship between science and technology.

CHI found a tripling in referencing from papers to patents between 1987/88 and 1993/94

(Narin et al, 1997). Narin interpreted this as evidence of a new paradigm of innovation in which emerging technologies tended to advance through direct translation of research into innovation as opposed to the more traditional model in which technology builds primarily on prior technological advance and the links to research were indirect. This striking result prompted bibliometricians and economists to suggest that instead of changes in innovation, the growth in referencing from patents to papers might be due to changes in the process of obtaining a patent. Three factors were put forward: a tightening in the rule of disclosure, increasing patent litigation and easier database access.

To examine these hypotheses, CHI investigated the patterns in referencing growth (Hicks et al., 2000). Because the three factors do not differentially affect patents in different technologies nor references of different types, the hypotheses predict growth rates would be equal across types of referencing (to U.S. patents, foreign patents and scientific journal articles) and across all technologies. If the rate of growth differs between technologies and between say, references to patents and to papers, then such broad explanations are inadequate. CHI found that growth in referencing to scientific literature far exceeded growth in referencing to US patents suggesting that tightening of the rule of disclosure cannot explain all the growth in referencing to scientific literature. Growth in referencing to U.S. patents does not vary much by technology in comparison to growth in referencing to scientific literature. This suggests that broad factors may well explain growth in referencing to U.S. patents but they cannot explain growth in referencing to scientific literature.

The idea that easier access to literature databases has fed the growth in referencing to papers is implausible on another level as well. Note that database accessibility should be of minor importance to applicants who, as experts in their fields, should have the literature relevant to their invention. It is unlikely that database accessibility increased during the 1980's and 1990's for U.S. patent examiners. Users of Dialog looking for scientific literature have had access to powerful searching capabilities since 1972. However, using Dialog is complicated, and in recent years the need to learn its complications has lessened as easier methods of searching have become available. But how relevant is this to U.S. patent examiners who work in very narrow technological "arts" and spend a great deal of their time finding prior art? In such a situation, a person who needed to use Dialog to access a literature source would need to do so frequently and could well have invested the time needed to learn how to do so. In addition, the patent office attempts to make available to examiners sources that are needed. For example, until 1995 in areas where foreign patents were important, foreign patents were obtained and classified into U.S. classes and made available to the examiners. Ultimately, easier database access seems unlikely to explain the growth in science referencing.

In summary, almost two decades of research have been undertaken to develop patent to paper references as indicators of science and technology linkage (Narin, 1976; Carpenter & Narin, 1978 and 1983; Carpenter et al. 1980; Carpenter, 1983; ABRC, 1986; Collins & Wyatt, 1988; Narin & Olivastro, 1992 & 1998; Narin et al. 1997, 1998). Although these indicators are obviously one tool among many, they nevertheless are a valuable tool with unique strengths, enabling quantitative comparisons of science-technology linkage across technologies, nations and time.

Elements of method

Defining paper and patent sets

The first step in any bibliometric evaluation is to define the set of papers or patents to be examined. There are two fundamentally different approaches to this, and the choice is made depending on the nature of the study. One approach is to compile a bibliography for authors, institutions or other units of analysis. The second approach is to compile a bibliography for a subject area.

Bibliographies for authors, institutions or nations are more or less difficult to compile depending on how clean the database is and whether a thesaurus is available. Author bibliographies require one to search for likely papers, incorporating likely misspellings, and to weed out papers belonging to other authors who share the same name. This is a time consuming but necessary first step. CV's or other bibliographies might seem like a good starting point, but they are very difficult to work with (Dietz et al., 2000). They also contain errors, missing information and are inconsistent – that is some people or institutions will be more thorough than others. Therefore, supplemental searching is required, inevitably such searches will find papers not on the bibliographies.

Subject area bibliographies present different challenges. Bibliographies of patents or papers in a subject area are compiled by developing a filter, or a protocol specifying the combination of addresses, words and combinations of words in the document, classifications, and citation relationships that are found through iterative experimentation to deliver a set of papers or patents that represents the subject of interest. The set will need to be weeded to eliminate irrelevant material. It is perfectly possible to work through this process using real intelligence, though the artificial variety is also applied to the task in the form of algorithms that cluster documents using one or more of the elements of a filter. Since it is important in evaluation to have a complete set of documents without irrelevant material, the results of algorithms should always be checked for completeness and accuracy.

Whole and fractional counting

How much has each author, institution or funding source contributed to the research underlying a paper? In an ideal world, it would be possible for all parties to reach agreement on their percentage contributions. In the real world, authors are likely to disagree – since each will most likely rate their own contribution as slightly more fundamental than those of their collaborators. Institutions would likely disagree with the authors, since they will take into account the infrastructure and facilities provided. And the funding agencies would have similar irreconcilable differences over whether, for example, the grant that brought a collaborator to the laboratory was more important than the support received by the Ph.D. student or the grant that bought the equipment.

In the bibliometric world, there are two sensible choices for allocating credit for papers: fractional and whole counting (one could also allocate credit to the first listed author, institution or funder, but that would be silly). In fractional counting each author/institution/funder is allocated a proportion of credit depending on how many are listed. So if there are two authors, each gets credited with ½ of a paper. In whole counting each author/institution/funder is allocated full credit for each paper in which they participated.

Fractional counting has the arithmetical advantage that everything adds to 100% with no double counting. If papers are whole counted, the numbers of papers cannot be added across author/institution/funders after the counting is done, in other words any aggregations have to be built into the original computations to avoid double counting collaborative work.

Fractional counting has the conceptual disadvantage of in effect penalizing people for collaborating. Since collaboration is generally considered a good thing these days and is encouraged by governments, any evaluation scheme which institutes what is in effect a penalty for collaborating must be suspect. One could take the point of view an economist might and say that if people are collaborating they must be doing less work and unless people can get at least twice as many citations by collaborating, they should not be doing it anyway – hence fractionation is not a penalty. Or, one could wonder whether people collaborate not to reduce effort, but rather to undertake research impossible otherwise. Collaborative papers are more highly cited than single author papers, though not enough to overcome the penalty imposed by fractional counting (Katz, 1997). Philosophically, one has to decide whether it is better for two authors to work alone and produce, for example, two papers cited five times or to collaborate on one paper that is cited 10 times. With fractionation, the authors get the same credit either way, though the contribution to knowledge is likely to be more significant with collaboration.

Institutional identification

In general, papers must be identified with institutions to be useful in evaluation. All author addresses are indexed in the *Science Citation Index* which is one of its unique strengths from the evaluation perspective. Unfortunately, addresses are not institutional affiliations, and a thesaurus of address variant unifications is needed to identify each paper with its institutions. Such a thesaurus forms a valuable part of an evaluation infrastructure as without one, data must be cleaned up for each study.

Funding sources

Many CHI evaluation studies include analysis of papers funded by an agency. This information is not included in databases but is compiled by CHI at the paper level for each study. There is no magic to this procedure; staff look up papers in the library and record the funding sources acknowledged by the authors. A certain percentage of papers will not be found - older papers are more likely to be missing than newer papers. (Therefore, in CHI's semi-routine lookup of funding acknowledgments for papers cited in patents an 11 year window is used - that is papers published up to 11 years before the patent is issued are looked up.) In addition, not all authors acknowledge funding. Authors at companies, or research institutes with intramural funding will not acknowledge funding; however, their funding can be inferred from their address. In addition, acknowledging varies by journal and field. Review papers are less likely to acknowledge funding, and as these papers can be quite highly cited, they should be handled with care in studies examining funding acknowledgments.

Citation counting

In a direct citation approach, the impact or quality of a set of research papers is ascertained by counting the numbers of citations the papers receive in the first three-to-five or more years after they are published. This requires the use of the *Science Citation Index* (SCI) since that is the only large database to include citation information, and requires that each citation be matched to the research paper that it references. Use of direct citation counting techniques requires that a sufficient number of years elapse for the papers to be cited, and that normalizations are available to account for the variation in rates of citation by subfield illustrated in Figure III-1. Direct citation counting allows one to identify high impact individuals, topics and research groups. Knowledge of the citation distributions by subfield and year enables one to characterize the citations to a group of papers as high, low or average relative to averages for subfields and for different combinations of citing and cited year.

The methodological discussion of strengths and weaknesses of bibliometric technique made clear that unless closely matched groups are being compared, normalized citation measures must be used. That is, a benchmark must be established against which to assess whether a citation count is high, medium or low given the publication year, journal, field, country etc. of the paper(s). There are several methods to establish such benchmarks

One commonly used method is to normalize by journal, that is to establish an expected rate of citation for each paper based on the average number of citations received by papers published in the same year and same journal. Using these norms, the number of citations accumulated to date by any paper can be compared against the expected value of citations for a paper in that journal and publication year. The trouble with this method is that an article in a weak journal that has accumulated five citations can often be considered more important than a paper published in a prestigious journal that has accumulated ten citations. There is an incentive for researchers to avoid publishing in *Nature* and *Science*. And an institution that improves the quality of journals in which it publishes will see a drop in its performance indicators. To avoid this sort of nonsense, a broader base for normalization is desirable.

CHI uses subfield-based normalizations, and there are several possible ways to do this. The most sophisticated, and earliest method used statistical theory to derive a standardized citation score. The basis of the standardized citation score was that rather than comparing papers using number of citations received, papers were compared using their relative positions in the citation distribution of all papers in the same subject area (McAllister et al. 1983).[37] This method was used in evaluations, but it fell out of use in part because the statistical calculations that stood between the citation counts and the score reduced the immediacy and comprehensibility of the results produced.

More frequently used was a direct comparison between the average citations per evaluated paper and the average citations per paper for the subfield as a whole. Here

[37] Strictly speaking, the distribution of all **U.S.** papers in the same subject area was used because U.S. papers were being evaluated. Of course, other national distributions or indeed the distribution of papers from a collection of countries could also be chosen as the basis for comparison.

citations were summed over some period of time following publication, the period of time being called a "window". Usually 4 or 5 year windows were used. In a frequently used variation, the share of evaluated papers found among the top 10%, or top decile, for their subfield was reported. Papers in the top decile are very highly cited papers, and for any set of evaluated papers, one would expect that 10% of them are among the most cited 10% in their subfield. Therefore, if 20% or 30% of the papers are among the top 10% in the subfield, that suggests quite strong performance. In the following chapter, there are examples of each of these normalization methods.

It is possible to argue that the subfield classification of journals is too crude a basis for normalization and finer distinctions should be drawn - for example, instead of environmental sciences, studies of pollution transport in water. We have already seen some of the problems here because in a sense journals would be a finer category for normalization, and the dangers of that approach have been mentioned. Finer distinctions are also subject to a problem we could label: "the goldfish bowl problem". For agencies, it is legitimate to ask: what is the standard that we are asking our researchers to achieve? The standard may be international, national or local. If the question addressed by an evaluation were - “how big is our fish?” - setting the standard means asking: do we want to have the biggest fish in the ocean, the biggest in the lake, or the biggest in the aquarium? The body of water can always be shrunk to the point that any fish looks large, but how informative is it to know that your fish is the largest in the goldfish bowl? Normalizations at the subfield level are a reasonable choice which does not demand that mathematicians achieve the impossible and attain citation rates comparable to biomedical researchers, but neither does it guarantee that every obscure mathematics paper will shine because it is compared against the four other most closely related papers, or because it was published in a generally low quality journal.

Journal Influence

Whereas citation counts are compiled for each paper, journal influence is a journal-based measure. The purpose of journal influence is to provide a proxy for citation counts that is available quickly, without waiting the four or five years needed to accumulate a reasonable citation history. The journal influence approach uses a score for each journal instead of exact citation counts to measure the impact of a set of papers. The journal influence score is, in essence, the weighted average number of times papers in a journal are cited, so that publication in a prestigious, highly cited journal is considered to be more influential than publication in a more obscure journal (Pinski & Narin, 1976). In the influence computation, each journal is assumed initially to have an influence weight of 1. The number of times that journal is cited by other journals is then divided by the number of references the journal gives out, to yield, as a first approximation to the influence weight of the journal, the ratio of citations received to references given. At the end of the first iteration, this influence weight is then used as part of the calculation to weight the references from that journal. That is, citations from a highly influential journal are considered to be more important than citations from an ordinary journal. The iteration is continued until a stable influence weight for each journal is obtained. That influence weight is divided by the number of papers in the journal, to yield an influence per paper. The net result is that influence per paper is essentially the weighted number of citations per paper that the average paper in the journal receives. For a large enough set of papers, say 20 to 50 or more, the influence per paper for that set of papers will be a good approximation to the long term citation performance of the papers. A more commonly used journal measure is the journal impact factor calculated by ISI. This measures average citations per paper for each journal (with some refinements), without the iterative element of the influence measure. (See Seglen, 1997 for warnings regarding using impact factors in evaluation.)

The major advantage of journal influence is that it is somewhat simpler to use, and it is available immediately. One does not have to wait three to five years to see how often individual papers are cited. In exchange for this timeliness, the identification of specific highly cited scientists or groups of scientists and highly cited papers is lost, and larger numbers of papers are required for statistical testing.

Recently the influence methodology has been adapted by computer scientists to rate web pages (Kleinberg, 1999). Their influence-related method of rating web pages has been incorporated into the Google search engine (www.google.com).

Subfields and fields

A second journal-level element used in the CHI studies reported here is the classification of papers into scientific subfields and fields based on the journal in which they were published. CHI's journal classification system places each of the journals in the SCI into one of approximately 150 subfields. These subfields are grouped into 9 fields. Although most journals fit reasonably well into a single subfield, approximately 100 journals, including Science and Nature, are so broad in scope that they are put into general fields. Individually classifying papers in these journals based on what fields they cite is a more advanced method of treating these journals (Glanzel, 1999).

Research level

The third journal-based classification parameter used in the CHI studies is research level. The idea behind research level was to develop a single number which would characterize an institution's publications on the basic to applied research spectrum. The inspiration came from the observation that applied journals tended to cite more basic journals, but the reverse was not true, that is basic journals did not cite applied journals. To implement the scheme, each of the SCI-covered journals was classified by CHI staff into one of four research levels from the most applied technology in "Level 1" to the most basic research in "Level 4". Table 3-3 illustrates the scheme with prototype journals for each of the four research levels. Note that a very basic physics journal such as the Physical Review is Level 4, whereas the Journal of Applied Physics is Level 3, the typical engineering journals are found in Level 2, and Level 1 is very applied, practitioner rather than research journals.

Fields differ in the percentage of papers within each level. Physics has virtually no Level 1 or 2 journals whereas the engineering subfields are mostly levels 1 and 2. Within some subfields, particularly some of the biomedical subfields such as cancer research, there is a wide spectrum of journals, from very clinical to very basic.

Table III-3. Research Levels and their Prototype Journals

Level	Name in both physical sciences and life sciences	Prototype Journal
1	Applied Technology	Tetsu To Hagane-Journal Of The Iron And Steel Institute Of Japan
	Clinical Observation	Gastrointestinal Endoscopy
2	Engineering/Technological Science	J Nuclear Science & Technology
	Clinical Mix	New England Journal of Medicine
3	Applied Research	J Applied Physics D – Applied Physics
	Clinical Investigation	Cancer Research
		J Clinical Investigation
4	Basic Scientific Research	Physical Review B – Condensed Matter
		J American Chemical Society
		J Biological Chemistry

Conclusion

Over the past several decades, Francis Narin and CHI Research have worked to develop the full potential of bibliometric indicators. Three types of linkage indicators have been developed: paper-paper, patent-patent and patent-paper. Each has its strengths and weaknesses and must be interpreted with care. The indicators also must be constructed with care. The databases from which the indicators are derived are not suitable for analytical work in their raw form. Thesauri, classifications and norms are needed to conduct analytical studies. The following chapter illustrates how bibliometrics using these analytical elements can be deployed to answer questions relevant to research evaluation.

Chapter IV. Examples of Evaluation Studies that Used Bibliometrics

by Diana Hicks, Peter Kroll, and Patrick Thomas

This chapter comprises short summaries of studies that used bibliometrics and were commissioned by government agencies to examine their programs.[38] There are four sections. The first describes three recent evaluative studies not conducted by CHI, but which contained bibliometrics. In addition, the first section contains a description of a recent quantitative case study by Bozeman et al. (the Research Value Mapping Project). The rest of the chapter describes work conducted by CHI over two decades for the U.S. Federal government and is divided into three parts covering: paper-paper, patent-patent and patent-paper bibliometrics. The summaries are presented in chronological order within each section. Most of these example studies are taken from CHI Research, Inc.'s archives, and in some cases the identity of the agency commissioning the work was removed. The intent of the chapter is to offer illustrations of the range of bibliometric work that has been conducted, the variety of ways in which the methods can be used and the variety of questions that can be addressed using the techniques.

The bibliometric methodology described in Chapter 3 serves as a reference for these summaries which often mention things like “journal influence” or “normalizations” which were explained in Chapter 3. In addition, the studies described here provide supporting material for Chapter 1 in that the early work for NIH which was quite widely known at the time is described and the recent build up of outcome related work is also visible in the patent-patent and patent-paper work.

[38] Ideally, each summary would have included a commentary on the reception and use of the study, but as is common, that information is not known to us in most cases.

List of Studies

A.

Examples of Recent Research Evaluations for the U.S. Federal Government that Included Quantitative Data

1. Evaluation of a National Funding Agency's Impact on Innovation

Roessner, D., B. Bozeman, I. Feller, C. Hill, N. Newman, *The Role of NSF's Support of Engineering in Enabling Technological Innovation*, first year final report for National Science Foundation (SRI International: Arlington VA, January 1997). Available at: http://www.sri.com/policy/stp/techin/

Roessner, D., R. Carr, I. Feller, M. McGeary, N. Newman, *The Role of NSF's Support of Engineering in Enabling Technological Innovation: Phase II*, final report to National Science Foundation (SRI International: Arlington VA, May 1998). Available at: http://www.sri.com/policy/stp/techin2/

Objective

To study how NSF support for engineering, especially research and related activities, contributed to the development and commercialization of recent, significant innovations. According to NSF:

> to conduct a systematic examination of the antecedent discoveries, events, people, interactions, and conditions that lead to the evolution of the 12 most significant engineering innovations to have emerged in the preceding decade to: (1) document NSF's involvement in bringing about the innovations; and (2) evaluate the significance of NSF's role in the broader context of the innovation's development.

Methodology

These studies used the retrospective case study approach to tracing the impact of scientific research on technology. This method was pioneered in the famous Hindsight and TRACES studies of the late 1960s and early 1970s that sought to identify the origins in science of significant innovations. The innovations studied were: Internet, magnetic resonance imaging (MRI), reaction injection molding (RIM), computer-aided design applied to electronic circuits (CAD/EC), optical fiber for telecommunications and analog cellular phone.

The method involved:

1. **Identifying** the technologies that underpinned each innovation, and deciding which were unique to the innovation and which were supporting technologies that already existed.

2. **Library search** of online databases to find all major works describing the development of the technologies.

3. **Bibliometrics -** Very small and tentative experiments in bibliometrics were performed using Research Front Databases of ISI. These databases use a clustering technique to group publications. The data were explored to show the close relationship of NMR imaging to NMR spectroscopy research, and to find NSF-funded researchers among major contributors in the field of composite materials. Patent databases were searched to find inventors, institutions, coinventors and citations of key research

literature. Two further experiments used co-citation analysis on papers and patents.

4. **Institutional Analysis -** Identify major organizations that developed the technologies from the library search, discussions with NSF staff, interviews and searches of NSF's awards database.

5. **Personal and Phone Interviews -** Contributors identified using the above techniques were interviewed about the history of the technology and NSF's role.

Results

In nearly all six cases, support for research and technology development by government, especially units of the Defense Department, played major roles. Without exception, the cases revealed the essential role that government support of education and training, especially graduate education, had on engineering innovation. Indeed, if there was a single, consistent pattern that stood out across all six cases, it was the critical role played by human capital in the form of individual inventors, technical entrepreneurs and students trained in the state-of-the-art who could continue to push technical advance. Regulatory policy shaped the course of innovation in the cellular phone and RIM cases. Given that successful innovations typically require several decades to evolve from conception to success in the marketplace, it is not surprising that fundamental research was found to play a supportive rather than central role in the six cases of engineering innovation.

NSF emerged consistently as a major, often the major, source of support for education and training of the Ph.D. scientists and engineers who went on to make major contributions to each innovation. NSF's support of university research infrastructure emerged as the likely second most influential activity. NSF provided major support for infrastructure in half the cases: internet, CAD/EC and MRI. NSF's direct research support was key to successful innovation in just one case: CAD/EC. NSF's research support produced knowledge essential to the evolution of all cases. NSF's organizational leadership was commanding, highly visible and unique in the internet case. In about half the cases, NSF shaped the evolution of research areas by encouraging university researchers to address problems relevant to industry using workshops and symposia which brought university and industry people together to discuss promising areas.

Bibliometric Results

RIM - 294 patents were found using keywords "reaction injection molding"; 64 were cited 5 or more times. Patents cited 7 or more times were virtually all related to RIM and were issued to private firms. The percentage of patents with references to scientific literature suggested the area was science linked when compared against Narin's published data.

NMR - ISI's research front database was used to produce a map of specialty clusters in the 1990 Research Front Database. The map showed two sets of interrelated specialties related to imaging and spectroscopy respectively. The strong link between the two areas confirmed findings from interviews. One tentative conclusion from this experiment was that the bibliometric mapping technique represents a potentially helpful tool for bounding the research themes relevant to an innovation.

Cellular telephony - papers - Five names were found as citing or cited authors in the 1988, 1989 and 1990 Research Front Database. Their number of papers in the Research Front Database was counted. One of the 46 clusters associated with the papers contained papers by a number of pioneers in the field suggesting that searches of this type would prove a valuable way of validating and extending the range of knowledge at the inception of a case study.

Cellular telephony - patents - 10 highly cited patents were identified in online searching. Patents cited by these 10 and cited by those and so on were retrieved. A co-citation analysis of these patents was conducted and a map was drawn. The map contained distortions as all multi-dimensional scaling maps do (the two most closely linked patents were not the closest together on the map). The map confirmed the finding that AT&T and Motorola were the only significant players developing the technology. Although crude and distorted, the map provided a useful overview of the patented technologies underlying cellular telephony, including an intimation of the digital technology to come. The technique merits further development and can contribute too understanding technologies relevant to the case studies.

Comment

The cases offer rich detail and a nuanced understanding of NSF's role in enabling technological innovation that can serve to enhance decision making and further research.

The bibliometrics did not get very far, seeming to produce what a filter would have done (see methodology chapter). The step of bounding the area of study, with a filter or a map, would ideally be the first step not the end point. Lack of normalizations hampered interpretation of data produced.

2. Evaluation of Funded Basic Research Outcomes

Barry Bozeman et al. - aka: The Research Value Mapping Project, "Qualitative-Quantitative Case Studies of Research Projects Funded by the Office of Basic Energy Sciences," Final report submitted to the Office of Basic Energy Science, Department of Energy, February 1999. Available at: http://rvm.pp.gatech.edu/papers/index.html.

Objective

To evaluate projects of the Office of Basic Energy Science, Department of Energy

Methodology

28 case studies and a survey of BES-funded projects were conducted over four years. Analytical models that track the flow of knowledge from research to outcomes were developed. The outcomes were modeled as a sequence of events in a branching structure. Hypotheses linking causal factors to outcomes were developed. Data was collected from interviews and documents. Indicators were developed for such cost and benefit variables as amount of funding, number of people, number of people trained, money earned, presence of technology transfer office, whether diffusion was planned, whether users needed to develop new manufacturing processes etc. Quantitative analysis was used to determine which variables were associated with successful outcomes.

Results

Stable BES funding is invaluable. BES funding was often seen as core funding by the researchers it supported and enabled the researchers to: build their capacity to produce important work, make contributions to knowledge and train students.

New organizational designs contribute to project effectiveness. BES projects exhibit a variety of innovative institutional and organizational designs, for example, large, fluid teams based on a set of scientific techniques or methods applicable to a range of fields; or new management infrastructures to support work ranging from fundamental science to marketable technology. The project manager holds these diverse enterprises together in way that seems more similar to industry R&D managers than to traditional principal investigators.

Basic research often produced applied results. Concrete evidence emerged not only of the malleability of basic research but also of its fertility for other aspects of the technical enterprise. If evaluation is focused narrowly on the goals stated at the beginning of the research, these wider impacts will be missed.

Interdisciplinary work requires different management than small-scale, individual investigator, disciplinary-oriented projects. BES projects are generally conducted by multidisciplinary, multifocus teams managed using traditional peer review combined with a nontraditional emphasis on institution- and capacity-building.

The results of university and federal laboratory projects were similar in range, technical focus and impact. Universities have access to graduate students but are hampered by disciplinary focus among colleagues. Government labs have better stability and clearer focus. Universities and federal laboratory researchers did not interact much.

Comment

The researchers emphasize the limits of evaluation methods that rely on the identification of projects. They find that "projects" are an administrative convenience used by agencies to manage funding but do not correspond to researchers' view of their work. They advocate taking a longer term view and focusing on the building of human capital and networks of people working together to achieve long-term technical goals. Without this longer term approach, they believe that the benefits of research will be underestimated. If RVM can develop a differentiation between productive and unproductive networks, their network evaluation approach would constitute an advanced and theoretically sound basis for evaluation of scientific research.

3. Text Mining to Identify Technical Areas Critical Infrastructure

Ronald N. Kostoff, "Implementation of Textual Data Mining in Government Organizations," presented at: Federal Data Mining Symposium and Exposition, Washington D.C. March 2000.

Background

Ronald Kostoff was for many years director of the Office of Technical Assessment at the U.S. Navy Office of Naval Research and in his work used quantitative evaluation techniques. His most recent work has been in applying techniques of text mining to management of science and technology. These techniques are often used in commercial competitive intelligence work, but are too micro level for most government evaluation studies.

Method

In 1998 Kostoff conducted a prototype implementation of a text mining approach. This involved several steps:

1) The iterative development of a filter[39] to identify papers associated with a technical theme (such as fullerenes, or ship hydrodynamics). Six databases were used as sources.

2) The frequency with which all words and phrases appeared in the documents was computed. Topical experts selected the useful phrases.

3) For each useful phrase, a dictionary of closely associated phrases was constructed by counting the number of times all other phrases occurred in close proximity to each useful phrase. Each associated phrase is assigned a measure of the strength of its association to the useful phrase. A threshold is used to filter out the most closely linked phrases. Topic experts identify the themes in the dictionaries and their conceptual relationships.

4) Analysis identifies pervasive technical themes in the database, the relationship among the themes and the relationship of supporting sub areas.

5) Bibliometric analysis is conducted using authors, journals, addresses etc. to relate the themes to performers.

Results

Bibliometrics identified the location of critical infrastructure in each technical area. This was useful for finding experts for workshops and review panels and for planning visits. Bibliometrics also allow productivity and impact to be tracked, and the critical intellectual heritage to be identified. Because no norms were available, it was

---

[39] See Methodology chapter for discussion of filters.

important to compare bibliometrics across disciplines so that anomalies in any one could be spotted and universal trends identified.

The phrases were classified by experts into a technical taxonomy containing categories useful to management. Phrases in Navy technical requirements documents were processed into the same taxonomy as the phrases in the papers. Quantitative estimates of levels of emphasis for each taxonomy category in both the requirements documents and papers were made. A matrix of requirements documents against papers allowed experts to judge the adequacy or deficiency of technical emphasis in each category. A similar, though softer process was undertaken with experts estimating the level of technical opportunity in each category. The resolution of the categories was an important parameter in the study. Finer categories ("welded titanium alloys" rather than "materials" for example) are more useful, but are more expensive and time consuming to construct.

The process centered on the subject experts, with the computer supporting their work. Therefore the conclusions reflect experts' biases and limitations. For a credible analysis that detects the maximum number of data anomalies, experts with diverse knowledge are required and a generalist is needed to identify unique patterns in a technical domain that the domain expert might not recognize as unique. From an organization's long-range strategic viewpoint, the main output is not the documents generated, but rather the broadening of the experts' perspectives. There was a steep learning curve for the experts, who had to learn how to use the tools to address the study's objectives and how to analyze and interpret the information produced.

Comment

Kostoff believes that all S&T management decision aids are inter-related and need to be integrated to support S&T strategic management. Thus, a program peer review should be accompanied by metrics to gauge progress toward strategic goals, should have roadmaps to place the program under review in its larger spatial and temporal context, should have text mining to insure roadmap comprehensiveness, and so on.

4. Benchmarking U.S. Research Fields

Experiments in International Benchmarking of US Research Fields
Committee on Science, Engineering, and Public Policy (COSEPUP)
National Academy Press, Washington D.C., 2000
Available at: http://books.nap.edu/html/exp_in_bench/pdf/

Objective

To experiment with a benchmarking methodology to evaluate the research-leadership status of the United States. International benchmarking compares the quality and impact of research in one country (or region) with world standards.

Methodology

Three areas were chosen: mathematics, immunology, and materials science & engineering. In each area COSEPUP appointed a panel of eminent scientists to produce a report. The methods used by the panels were:

- The Virtual Congress - panel members called leading experts in sub-subfields to ask who are the 5-20 best people in the world. The resulting lists of people were augmented with country affiliations to produce the indicator.
- Citation Analysis - An existing British analysis was used in each field. The immunology panel bought a "high impact" immunology database from ISI and prolific authors were ranked by citation counts and classified by country.
- Journal Publication Analysis - Five journals were scanned and the locations of principal investigators and their subfields were tabulated.
- Quantitative Data Analysis - In general, data on education and funding were not available internationally for comparison.
- Prize Analysis - In each field, the number of US and non-US recipients of key prizes were tallied, keeping in mind that recipients may have moved.
- International Congress Speakers - US representation among conference speakers was tallied, with an awareness that conference organizers strive for geographic balance by inviting speakers from a range of countries.

Results

Each panel concluded that the United States was at least among the world leaders in its field. However, each panel also identified subfields in which the United States lagged the world leaders. Each panel identified key infrastructure concerns.

Comment

The report contains interesting tensions. For example, it denounces quantitative indicators, which is a point made routinely in National Academy publications. The report states that quantitative indicators are useful, but by themselves inadequate, because, and this is a full recount of the reasoning relevant to bibliometrics, "for example a paper that describes truly innovative research may receive few citations if no on else is doing comparable work." Therefore, "expert judgment of panel members afforded the most effective means for assessing research" (pp. 6-7). Despite this, when the work had to be done, the scientists on the panels used bibliometric information.

The report also emphasizes cost effectiveness, which is an important element in a method proposed for routine use. A number of very senior scientists both domestic and foreign were used in producing the report, and the nature of the material meant that they had to make a greater commitment to writing the report than is common in Academy committee work. Yet, they were not paid for their time. Thus the cost of their participation was unwittingly borne by their institutions and the agencies awarding them their grants (if the method were used on a routine basis, foreign governments would be subsidizing U.S. science policy making). Thus, the procedure was very expensive but the true costs were understated by including only direct billings to the Academy. See Kostoff's handbook for a calculation of full costs for studies of this type.

Each benchmarking group was charged with assessing whether the U.S. is the world leader in a field of science. The report recognizes and wrestles with the inherent subjectivity involved in asking very senior American scientists to decide whether their life's work has led to American leadership in their field. The scientists involved face a difficult choice: if they say the U.S. is a world leader, perhaps some might conclude that they do not need more money. Yet it would be embarrassing to say that the U.S. was not a world leader, not least because the U.S. effort is so much larger than anyone else's.

B.

Paper-to-Paper Bibliometrics:

Evaluating Research Outputs

1. Evaluation of the Effects of Legislation Seeking to Foster Targeted Research

Samuel Reisher and Francis Narin, May 1980

Objective

Quantitative indicators were developed to examine the effect on research activity of the National Heart, Blood Vessel, Lung and Blood Act of 1972. The 1972 Act gave the National Heart, Lung and Blood Institute (NHLBI) broad authority to combat heart, lung, blood vessel and blood diseases. The analysis focused on research in three areas: structure and function of the lung (LSF), chronic obstructive pulmonary diseases (COPD), and hypertension (HPT). LSF and COPD were specifically targeted by the 1972 Act, whereas HPT had been targeted by earlier legislation. Therefore, HPT acted as a kind of control against which to assess trends in COPD and LSF research.

Methodology

The bibliography for this study was built using three databases, each of which contributed unique information on the papers. In the end, CHI compiled a database containing the following information: bibliographic reference, institutions, funding acknowledgements, citation counts, research level, journal influence, subject area. Medline subject classifications were used to identify papers in the three research areas. Working with the client, sophisticated definitions of each research area were developed and each of about 15,000 papers were classified. Since the topic of each paper received such detailed attention, an even more detailed classification of papers was developed and a content analysis was undertaken.

Results

After the 1972 Act, the share of papers in the two targeted areas acknowledging NHLBI support increased. This effect was not seen in the control area. Also, after the 1972 Act, NHLBI papers in the two targeted areas increased their average citation rate, and the likelihood that they would be found among the most highly cited publications. This was not seen for the control area.

The study found that the 1972 Act came at a time of decreasing support for biomedical research. The number of biomedical papers in many areas declined around this time, and the size of the targeted fields (COPD and LSF) did not increase. HPT had been targeted by earlier legislation and was growing before the 1972 Act and continued to grow afterwards.

Other characteristics of the papers were unaffected by the 1972 Act. During this time, the papers examined shifted towards more basic biomedical research, but both targeted and control area papers exhibited this trend. Content analysis also revealed (in both target and control areas) an increased emphasis on the identification and characterization of biologically active substances in which NHLBI funded work was leading the way. This was a crucial shift to more scientific type of biomedical research in which NIH in general led.

Comments

Here bibliometrics was used to identify the effects of a piece of research-related legislation. Quite subtle effects were established.

(CH 051)

2. A Comprehensive Benchmarking of a Research Laboratory's Published Output

Paul McAllister and Francis Narin, October 1982

Objective

To develop quantitative indicators of the scientific performance of Navy research laboratories using the open published literature.

Methodology

500 individual publishing units were identified including individual RDT&E facilities of the US military. In addition 19 U.S. non-military federal institutions, 18 Federally Funded Research and Development Centers and 30 companies were separately identified and studied in the analysis. The data included essentially all U.S. non-university papers. A three-part characterization of each publishing unit was developed encompassing the publication size, subject emphasis and citation merit of each unit's research. These characterizations were devised so that they were completely comparable across units engaged in very different fields of inquiry.

Results

The premier role of the Navy Research Laboratories (NRL) among all Department of Defense (DOD) facilities was clear. It was the single largest DOD publisher, on a par with major federal laboratories such as Oak Ridge or Argonne. NRL scientists are active in physics, earth & space science and engineering. NRL work was also quite highly cited, with the highest overall citation score of any Navy facility - oceanography was particularly notable in this regard. Several subfields in which NRL citation performance lagged were identified. Navy papers were more highly cited than Air Force or Army papers and NRL was largely responsible for this. The top 25 most highly cited papers in each of nine major fields in each of two years was determined (450 papers), and Navy authors participated in 18 of these - see attached table IV-1. DOD scientists in general were cited less frequently than would be expected, which might have been attributable to the specialized nature of their research.

Comment

This study applied bibliometrics to examine the publishing component of output from laboratories whose main mission was not publishing. It is notable for the breadth institutions brought into the benchmarking effort.

(CH185)

Table IV-1. Navy Papers in the Top 25 Most Highly Cited in Their Field

A. 1974 PAPERS

Field	Rank in the Field	Number of Citations	Journal	Authors	Title	Corporate Address
Biomed Res	21	82	Science	Nelsonre. WA Flanderm. RR Hawthorn. PK	Banded Marker Chromosomos as Indicators of Intra-species Cellular Contamination	USN, Biomed Res Lab Univ Calif, Sch Publ Hlth, Cell Cu
Earth & Space	12	50	J Geoph Res	Vogt. PR Avery. OE	Detailed Magnetic Surveys in Northeast Atlantic and Labrador Sea	USN, Oceanogr Off
Eng & Tech	5	35	P IEEE	Taylor. HF Yariv. A	Guided Wave Optics	USN, Electr Lab Ctr Caltech, Dept Elect Engn
	7	31	Computer Ph	Manning. I Mueller. GP	Depth Distribution of Energy Deposition by Ion-Bombardment	USN, Nucl Sci Div, Res Lab
	22	22	IEEE MICR T	Granatst. VL Herndon. M Parker. RK Schlesin. SP	Strong Submillimeter Radiation from Intense Relativistic Electron Beams	USN, Res Lab
Psychology	8	16	Psychophysl	Johnson. LC Naitoh. P Moses. JM Lubin. A	Interaction of REM Deprivation with Total Sleep Loss-Experiment 2	USN Hosp Neuropsychiat Res Unit
	9	15	Psychophysl	Lubin. A Moses. JM Johnson. LC Naitoh. P	Recuperative Effects of REM-Sleep and Stage 4 Sleep on Human Performance After Complete Sleep Loss-Experiment 1	USN Med Neuropsychiat Res Unit
	23	11	Pharm Bio B	Shearer. DE Fleming. DE Bigler. DE Wilson. CD	Suppression of Photically Evoked After Discharge Bursting Following Administration of Anti-convulsants in Waking Rats	USN, Undersea Ctr Vet Admin Hosp, Neuropsychol Res Lab Brigham Young Univ, Dept Psychol
Mathematics	9	6	Am Math Mo	Gaskell. RE Klamkin. MS	Industrial Mathematician Views His Profession-Report of Committee on Corporate Numbers	Ford Motor Co., Scientific Res Sta USN, Post grad Sch, Dept Math
	16	5	Math Comput	Neild. C	3-Rank of Quadratic Fields and Euler Product	USN, Ship Res & Dev Ctr, Comp & Ma

3. Comparison of Funding Mechanisms: Centers vs Peer Reviewed Grants

Samuel R. Reisher and Francis Narin, July 1984

Objective

This study evaluated a program of funding centers by comparing the centers' scientific output to the output of scientists funded by traditional investigator-initiated grants. The study sought to assess whether the centers program was achieving its stated goals and whether there was any compromise in scientific quality as compared with the output from peer reviewed grants.

Methodology

A database of papers published by the centers and by principal investigators at the centers and recipients of R01 grants was painfully pieced together from information provided by the client combined with two publication databases. CHI added to this citation and institution data from the *Science Citation Index*, subfield classification, classification of research level basic to applied, and support acknowledgements. The various combinations of center papers, papers by center PI's, and papers by R01 recipients produced seven mutually exclusive categories of papers. The bibliometric characteristics of these categories were analyzed in relation to center goals.

Results

The centers were found to have attracted scientists with broad research backgrounds and interests in basic biomedical research. The centers scientists attracted additional funding sources including support from other parts of NIH. After joining a center, scientists continue to publish papers in a broad range of subfields. Centers fostered increased collaboration among individuals. Papers of center scientists are cited by papers from outside the center's focus. Center scientists are as productive and their papers are as highly cited as those of R01 PI's. See Table IV-2 which relates goals to bibliometric characteristics and evaluation results.

Comments

This study is a good example of using bibliometrics to establish quantitatively whether a program is reaching its pre-stated goals. It demonstrates that a variety of goals can be evaluated bibliometrically.

(CH183)

Table IV-2. Policy goals, bibliometric measures & results in Center-R01 study

Center program goal	Bibliometric measures	Results
	research level	• Center papers were much more oriented to basic research than R01 supported papers.
To attract basic biomedical scientists to develop a more interdisciplinary approach to center research.	subfields	• Center: 30% of papers appear in biochemistry & molecular biology subfield and 20% in center focus. R01: 50-60% of papers in center focus, 10% in biochemistry & molecular biology. • New principal investigators (PI's) in centers more often published primarily in fields other than center focus as compared to new R01 P.I.'s
	citing papers	• R01 papers were more likely to be cited by center focus papers and by papers acknowledging NIDR support. Center papers were more likely to be cited by papers produced at other centers.
Attract new types of money to dental research	support acknowledgements	• Center papers acknowledged about twice as many sources of support as R01 papers
Attract new people to dental research	individual's publishing histories	• Centers recruited more scientists who published in the basic biomedical sciences. • Centers & R01 were equally effective in recruiting new scientists, those without publications in the previous 6 years.
Encourage collaboration	number of co-authors	• Center papers averaged 3.1 authors; R01 2.9 authors. Papers with 4 or more authors comprised 33% of center papers and 26% of R01 papers.

	collaborating institutions	• Both averaged 1.3 institutions per paper
Produce research of as high a quality as produced when each project is individually peer-reviewed (R01 mechanism)	number of papers	• Center and R01 scientists similar. Those who have both center and R01 support publish about 50% more papers per year than those with either center or R01 funding.
	number of citations (controlled for year and subfield)	• Center and R01 scientists similar. Those who have both center and R01 support are more highly cited than those with either center or R01 funding.

(CH183)

4. An Example of Routine Bibliometric Profiling of Research Institutes

Francis Narin & CHI staff (undated)

Objective

An analytical characterization of papers supported by the National Institute of Child Health and Human Development. Parallel reports were prepared for each part of the NIH.

Methodology

As part of an ongoing effort, this report drew on a specialized bibliometric database CHI constructed for NIH which contained information on funding acknowledgements for papers in leading biomedical journals. This was combined with citation and institutional information from the *Science Citation Index*. For these reports, CHI compiled an analytical database containing the following information on each paper: bibliographic reference, institutions, funding acknowledgements, citation counts, research level, journal influence, subject area. In addition, citation links were used for a cross-citing analysis.

Results

NICHD supported a fairly constant mix of clinical and basic research over the period. The number of NICHD papers declined by about 3%. The Institute's major impact was in the subfields of Endocrinology, Obstetrics and Gynecology, Pediatrics, Anatomy and Morphology and Embryology. NICHD accounted for 12% of the papers in Embryology during 1973-80 and 10% of those in Anatomy and Morphology. Although support of Biochemistry and Molecular Biology was emphasized (17% of NICHD-sponsored papers) the Institute accounts for only 1.6% of the papers in that subfield.

NICHD's "activity index" - a measure of the Institute's publications in a subfield in relation to all publications in that subfield - was highest for Embryology: 7.2 in 1977-80, or 7.2 times the "expected" level. The institute supported more papers than any other NIH institute in the subfields of Fertility, Obstetrics-Gynecology, Pediatrics, Anatomy-Morphology and Embryology.

Subfields exhibiting large changes in NICHD supported papers were: Nutrition and Dietetics - 33% increase; Cell Biology/Cytology/Histology - 31% increase; and Fertility - 19% decrease. Research in Geriatrics declined with the creation of the National Institute on Aging in 1974.

NICHD's record of supporting outstanding investigators is manifest in citations to the Institute's publications. In 15 of the 16 subfields in which the Institute was most active, 10% or more of the Institute's papers were in the top decile of cited papers. This indicates that the quality of the Institute's programs has risen or been maintained despite funding restrictions in place at the time.

That NIH Institutes are interdependent in their research is clear from interaction in citation. Only 20% of references in NICHD supported papers were to research supported by that Institute; the rest were to papers supported by others. Conversely, 84% of the citations to NICHD research were in papers with other support. Interdependence was strong among NICHD, NIADDK, NCI and NIGMS.

Several figures and tables are attached for illustrative purposes. See figures IV-1 and IV-2 and the publication profile that summarizes the institute's bibliometric activity - Table IV-3 pages 1& 2.

Comments

This report is an example of bibliometrics used on an ongoing basis to provide comparable institutional profiles. It is probably not coincidental that NIH funding is relatively secure and their institutes operate at the highest levels of scientific excellence. Therefore NIH would have little fear of weaknesses being exposed through quantitative measures.

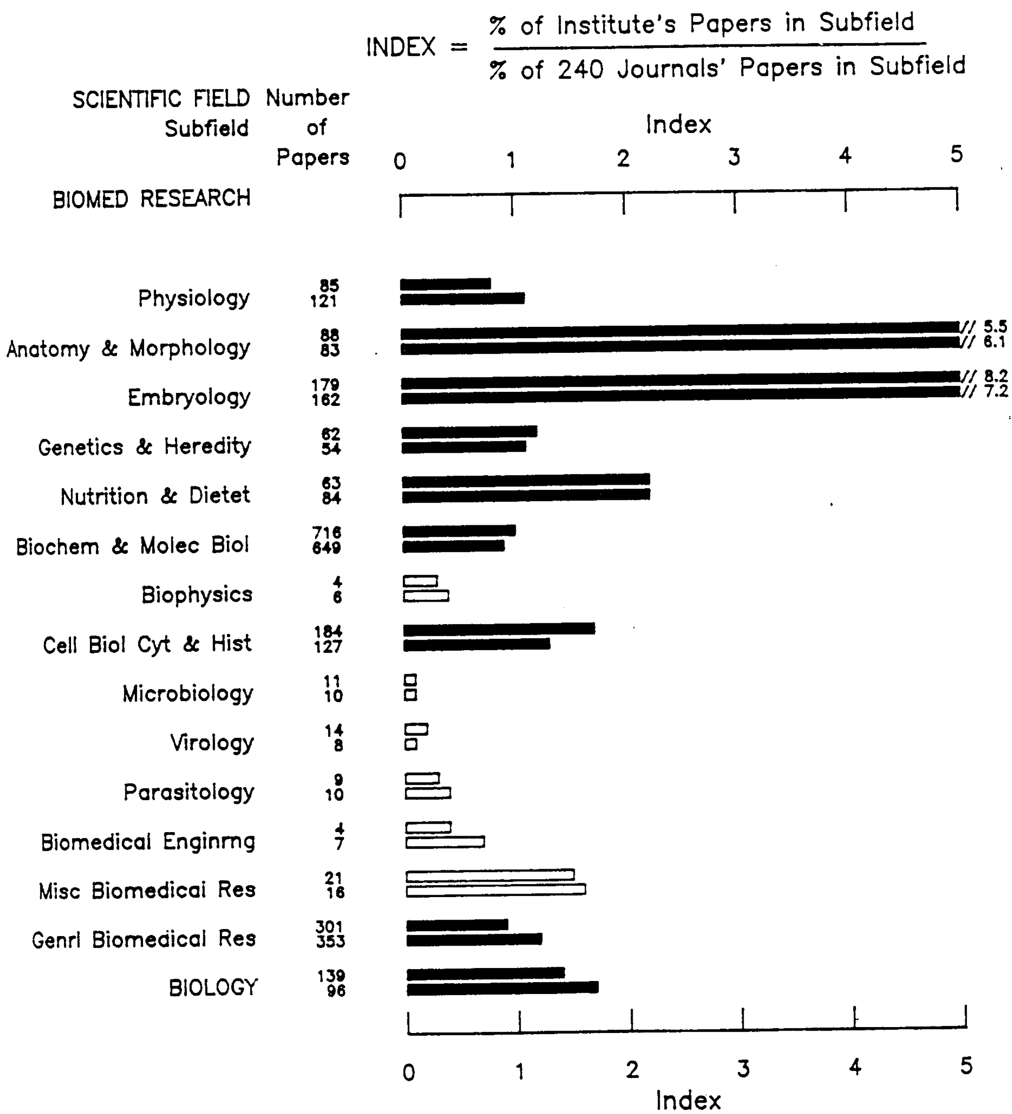
(CH237)

Figure IV-1.

NICHD Activity Indexes in Subfields of Biomedical Research

240 BID—MEDLINE Journals

1973—1976 over 1977—1980



Note: Open bars designate subfields that contained less than 1 percent of the papers acknowledging this Institute's support in the given period.

Figure IV-2.

Universe of Publications Cited by and Citing NICHD-Supported Papers

Area of Circles Proportional to Number of

Publications Supported; Citations by Percent

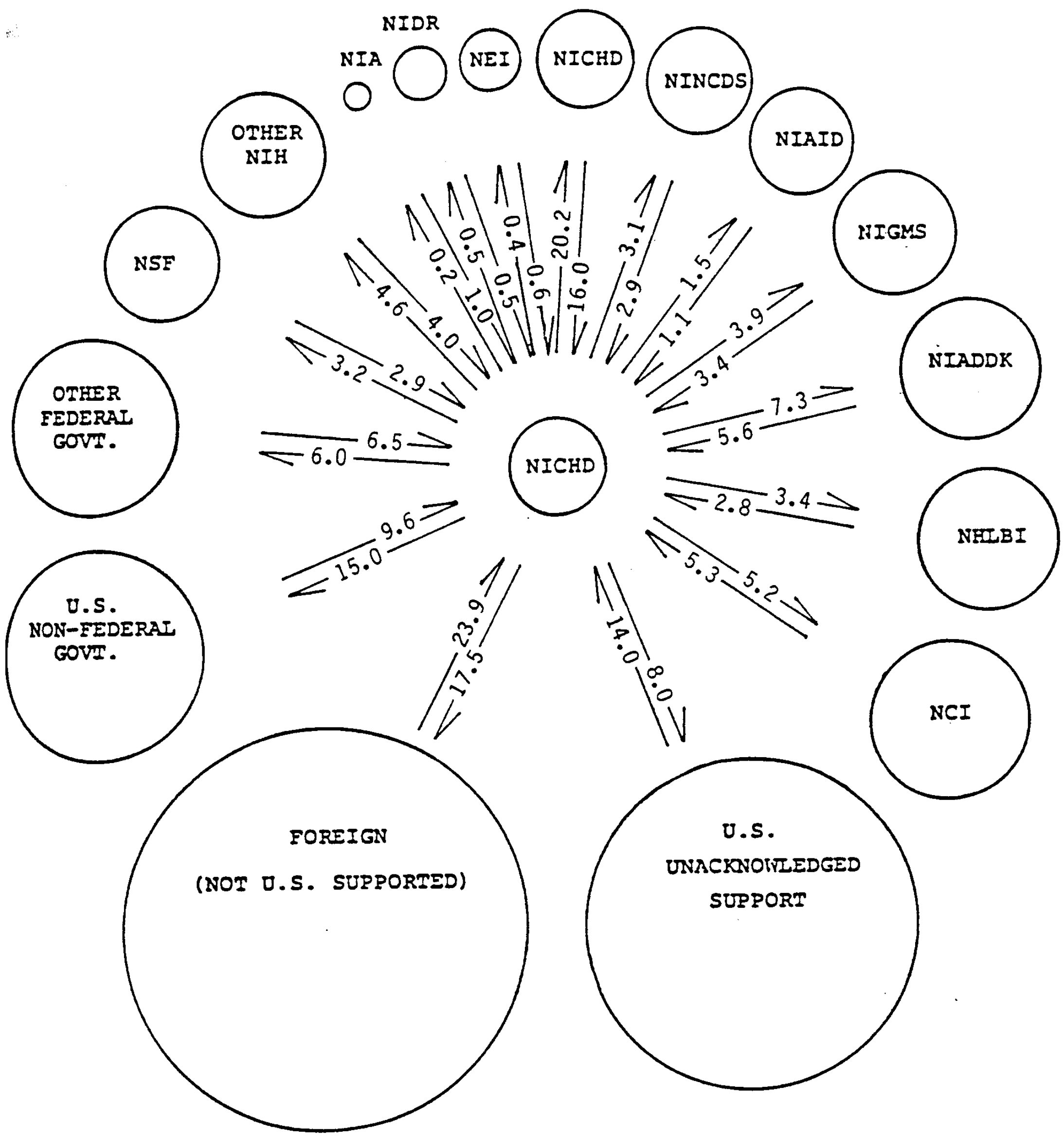


Table IV-3.

Bibliometric Profile of NICHD for the Cited Years 1977-80

Citation Counts Scaled to 1973 Equivalents

Based on Cites Received from All Citing Years, 1973-80*

FIELD/LEVEL		NUMBER PAPERS	ACTIV INDEX	% INT EFFRT	% EXT EFFRT	TOTAL CITES	CITES PER PAPER	CITE T-SCR	% PAPERS AMONG TOP 10%
****************** NORMALIZED AGAINST FIELDS [1-2] ******************									
FIELDS [1-2]		3892.	1.00	100.00	1.70	23929.	12.0	52.01	12.41
CLIN MEDICINE	[1]	2203.	0.93	56.60	1.57	11387.	10.3	53.63	16.25
BIOMED RESEARCH	[2]	1689.	1.12	43.40	1.90	12477.	14.0	50.46	10.75
LEVEL									
CLINICAL-MIX	[1-2]	773.	0.59	19.85	1.00	3512.	8.5	53.71	18.96
CLIN INVESTIGATN	[3]	1336.	1.39	34.32	2.35	7098.	11.0	52.54	12.79
BASIC RESEARCH	[4]	1784.	1.10	45.83	1.87	13196.	14.1	50.42	10.71
****************** NORMALIZED AGAINST FIELDS [1-9] ******************									
FIELDS [1-9]		4226.	1.00	100.00	1.77	24407.	11.2	51.68	12.68
CLIN MEDICINE	[1]	2203.	0.89	52.13	1.57	11387.	10.3	53.63	16.25
BIOMED RESEARCH	[2]	1689.	1.07	39.97	1.90	12477.	14.0	50.46	10.75
BIOLOGY	[3]	96.	1.73	2.26	3.06	310.	5.7	54.62	22.87
CHEMISTRY	[4]	18.	0.36	0.42	0.64	87.	9.1	52.34	3.51
PHYSICS	[5]	13.	1.78	0.31	3.15	40.	6.3	53.58	21.05
EARTH & SPACE SC	[6]	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
ENGRNG & TECHNOL	[7]	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
PSYCHOLOGY	[8]	208.	3.29	4.91	5.81	492.	4.2	54.26	18.31
MATHEMATICS	[9]	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00

NOTE: CHEMISTRY AND PSYCHOLOGY DATA SHOULD BE TREATED CAUTIOUSLY OVER TIME DUE TO INCONSISTENT DATABASE COVERAGE.

*1979-80 papers not included in present citation data.

Table IV-3. (Continued)

Bibliometric Profile of NICHD for the Cited Years 1977-80

Citation Counts Scaled to 1973 Equivalents

Based on Cites Received from All Citing Years, 1973-80*

(Subfields Normalized Against Fields 1-2)

FIELD/SUBFIELD	NUMBER PAPERS	ACTIV INDEX	% INT EFFRT	% EXT EFFRT	TOTAL CITES	CITES PER PAPER	CITE T-SCR	% PAPERS AMONG TOP 10%
CLIN MEDICINE [1]								
GENRL & INTERNAL MED	221.	0.68	5.67	1.15	1595.	13.4	53.29	12.78
ALLERGY	3.	0.31	0.06	0.53	2.	2.6	38.12	0.00
ANESTHESIOLOGY	7.	0.31	0.18	0.53	14.	2.8	47.01	20.00
CANCER	37.	0.23	0.95	0.39	304.	13.6	52.54	9.70
CARDIOVASCULAR SYSTM	8.	0.08	0.20	0.13	33.	6.9	43.61	0.00
DENTISTRY	4.	0.07	0.11	0.11	1.	0.4	44.20	0.00
DERMAT & VENERL DIS	7.	0.24	0.17	0.40	22.	5.8	55.89	21.74
ENDOCRINOLOGY	622.	5.48	15.99	9.29	4284.	15.5	58.57	14.94
FERTILITY	160.	5.15	4.12	8.73	483.	5.7	52.01	18.58
GASTROENTEROLOGY	7.	0.20	0.17	0.33	23.	8.8	54.63	12.50
GERIATRICS	12.	0.65	0.32	1.10	20.	2.7	54.49	25.58
HEMATOLOGY	7.	0.16	0.18	0.27	12.	5.6	45.31	0.00
IMMUNOLOGY	91.	0.51	2.35	0.87	572.	10.8	46.24	6.58
OBSTETRICS & GYNECOL	206.	3.57	5.29	6.06	837.	8.3	55.27	24.09
NEUROL & NEUROSURG	241.	0.98	6.19	1.67	1248.	10.2	50.17	9.44
OPHTHALMOLOGY	9.	0.17	0.22	0.28	72.	14.0	68.94	35.48
ORTHOPEDICS	3.	0.12	0.08	0.20	0.	0.0	44.64	0.00
ARTHRITIS & RHEUMAT	3.	0.18	0.08	0.30	3.	2.1	42.71	0.00
OTORHINOLARYNGOLOGY	17.	0.51	0.45	0.87	61.	4.2	54.12	24.14
PATHOLOGY	20.	0.39	0.51	0.66	69.	7.4	54.11	25.00
PEDIATRICS	315.	3.52	8.09	5.97	1148.	7.4	54.60	19.89
PHARMACOLOGY	77.	0.49	1.97	0.82	365.	9.6	50.91	14.10
PHARMACY	25.	0.47	0.64	0.80	102.	6.4	53.92	20.83
PSYCHIATRY	14.	0.24	0.36	0.41	22.	3.1	46.13	9.30
RADIOLOGY & NUCL MED	8.	0.08	0.20	0.13	14.	4.3	49.53	15.00
RESPIRATORY SYSTEM	10.	0.47	0.27	0.80	60.	9.2	67.48	30.77
SURGERY	8.	0.05	0.20	0.08	8.	2.6	45.40	0.00
TROPICAL MEDICINE	2.	0.13	0.04	0.21	0.	0.0	0.00	0.00
UROLOGY	13.	0.30	0.32	0.51	35.	4.2	52.51	18.00
NEPHROLOGY	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
VETERINARY MEDICINE	16.	0.33	0.40	0.56	55.	6.5	63.36	35.29
ADDICTIVE DISEASES	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
HYGIENE & PUBL HLTH	32.	1.21	0.83	2.05	76.	4.1	53.72	19.09
BIOMED RESEARCH [2]								
PHYSIOLOGY	121.	1.11	3.11	1.88	833.	14.7	51.35	9.97
ANATOMY & MORPHOLOGY	83.	6.06	2.13	10.28	320.	6.9	54.27	19.49
EMBRYOLOGY	162.	7.21	4.17	12.23	1004.	12.0	57.87	16.73
GENETICS & HEREDITY	54.	1.14	1.38	1.94	188.	6.4	45.57	7.95
NUTRITION & DIETET	84.	2.20	2.16	3.73	302.	6.8	50.48	8.24
BIOCHEM & MOLEC BIOL	649.	0.94	16.67	1.60	5419.	15.4	50.43	12.84
BIOPHYSICS	6.	0.42	0.16	0.71	18.	5.9	48.87	0.00
CELL BIOL CYT & HIST	127.	1.25	3.25	2.12	1047.	14.9	53.23	10.21
MICROBIOLOGY	10.	0.13	0.25	0.23	49.	10.9	57.64	29.63
VIROLOGY	8.	0.11	0.21	0.19	83.	20.7	65.70	25.00
PARASITOLOGY	10.	0.41	0.25	0.70	35.	6.7	61.31	25.81
BIOMEDICAL ENGINRNG	7.	0.67	0.18	1.14	12.	2.9	54.28	24.00
MISC BIOMEDICAL RES	16.	1.56	0.42	2.64	18.	1.8	49.99	8.20
GENRL BIOMEDICAL RES	353.	1.22	9.06	2.08	3280.	18.2	51.59	12.36

5. An Assessment of the Factors Affecting Critical Cancer Research Findings

Francis Narin, September, 1987

Objective

The National Cancer Institute (NCI) is responsible for conducting and supporting research related to the cause, prevention, diagnosis and treatment of cancer. The purpose of this study was to determine who were the principal contributors to selected advances in cancer research in the period 1965-1982, how they were supported (contract vs grants vs center support), and where they performed their work so as to determine if a significant relationship exists between the research event, the funding mechanism, and the location of the performer.

Methodology

An expert panel selected 13 important advances in cancer research and identified the major research events underlying each advance. The research events were arranged chronologically and thematically to produce a historiographical trace of each event - see Figure IV-3 attached. Each trace was associated with a set of key documents, 90% of which were papers, representing important advances in cancer research. The expert panel reviewed and commented on the draft traces which were revised and finalized. Four sets of papers were developed: 1) papers on the traces themselves, 2) closely related, contemporaneous papers (co-cited papers), 3) precursor papers cited by the trace papers, 4) less closely related, contemporaneous papers. Each paper was characterized by: support source (NCI, other NIH, other US etc.), mechanism (grant, contract, center etc.), research location (Harvard, NCI intramural etc.) and number of times cited.

Results

The NCI-supported papers were more heavily represented among the traces papers than in the surrounding sets representing a broader range of cancer research. Each of the major support mechanisms used by NCI contributed to important cancer research findings and the surrounding papers in equal measure. All the papers on the traces were extremely highly cited for their subfield and year of publication. Important research advances came from both large and small research institutions, U.S. and foreign institutions, universities and medical schools, as well as from the NCI itself. Thus, support mechanisms and institutional settings were not key factors affecting research advance, and by implication the procedures used to screen research proposals at NCI were choosing high impact research regardless of the support mechanisms and institutional settings.

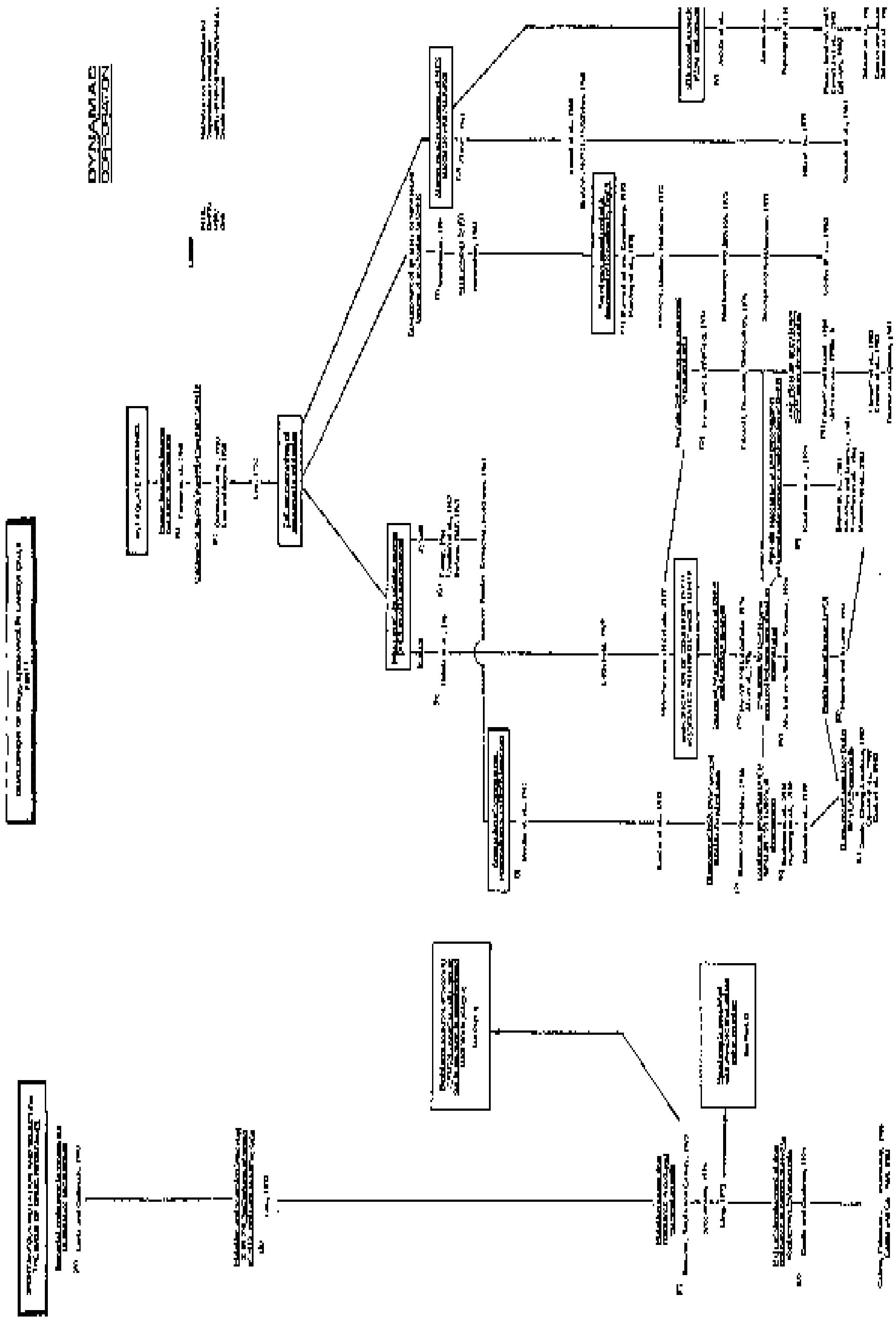
Comment

This study is a unique combination of tracing and bibliometric methods. Stronger results might have been produced if more norming had been used. That is, the share of NCI funding to the various mechanisms could have been compared against the share of papers reporting each type of funding, and the publishing size of the institutions could have been compared to the number of papers appearing in the traces.

(CH240)

Figure IV-3.

TRACE #2: Development of Drug Resistance in Cancer



6. Comparing Intramural and Extramural Support Mechanisms

James Corrigan and Francis Narin, October 1989

Objective

To compare the characteristics of NIH intramural and extramural work.

Methodology

As part of an ongoing effort, this report drew on a specialized bibliometric database CHI constructed for NIH which contained information on funding acknowledgements for papers in leading biomedical journals. This was combined with citation and institutional information from the *Science Citation Index*. For these reports, CHI compiled an analytical database containing the following information on each paper: bibliographic reference, institutions, funding acknowledgements, citation counts, research level, journal influence, subject area.

Results

Evidence of a shift towards more scientific medical research in NIH related papers was found in an increasing emphasis on the more basic subfields of Biomedical Research at the expense of Clinical Medicine. Papers in the field of Clinical Medicine tended to show higher relative citation performance than those in Biomedical Research.

As a percentage of the database, extramural papers were 15% and intramural 1.5%. Mean cites per paper for both extramural and intramural papers were well above database averages, receiving, respectively, 1.7 and 2.5 times as many citations as all papers. 20% of extramural and 30% of intramural papers were found among the top 10% most cited papers in their subfields (see attached Figure IV-4 for details). A consistent rank ordering on the various citation performance measures was observed in which: NIH intramural papers > NIH extramural papers > all U.S. papers > all papers in the database.

Comment

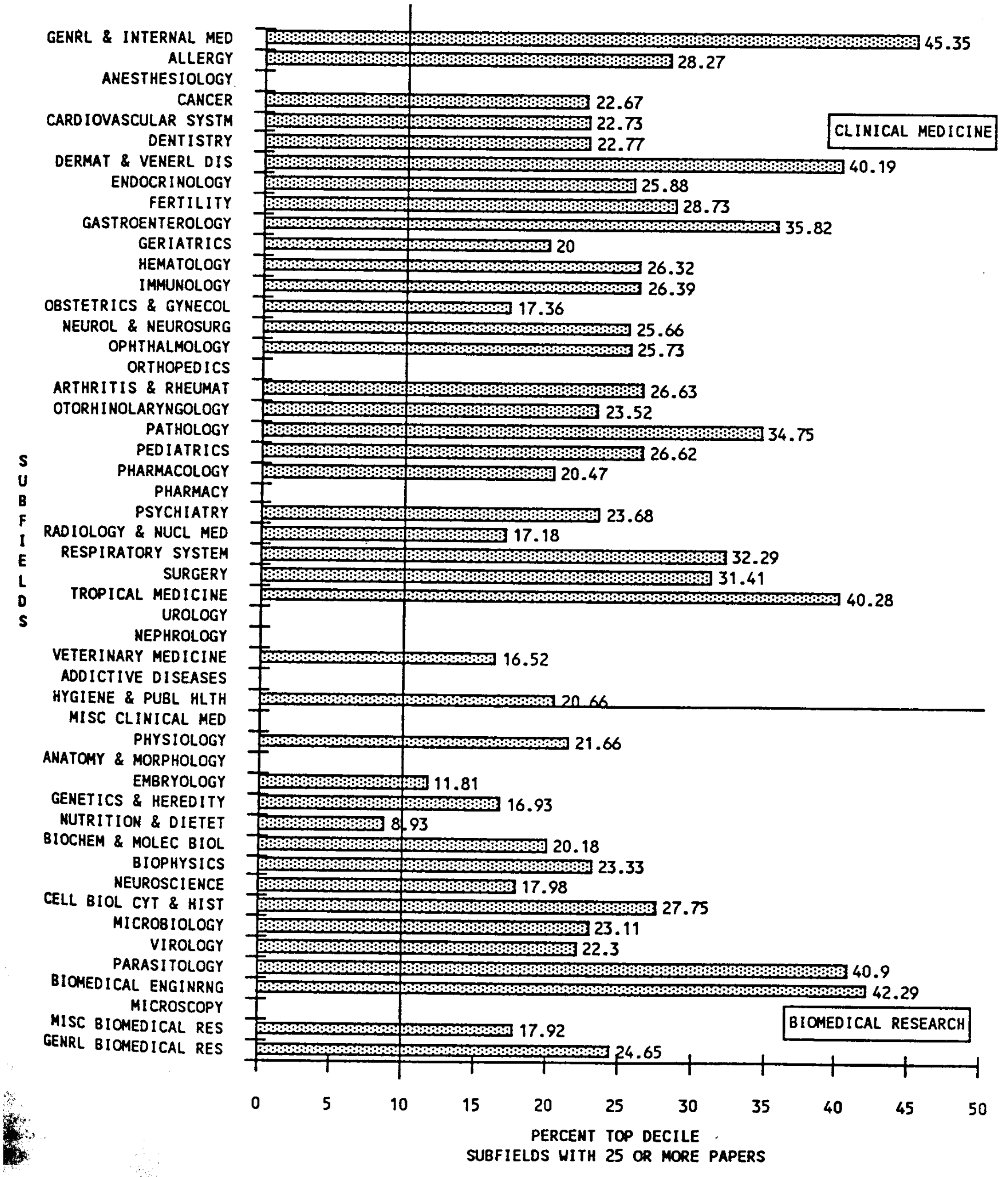
Here two support mechanisms were evaluated. Intramural means NIH's own laboratories and extramural means NIH money awarded to researchers in universities through peer-reviewed grants. This was part of an ongoing effort during which specialized databases were constructed and then put to use examining specific questions as well as producing regular reports on NIH institutes.

(CH275)

Figure IV-4.

All NIH International Percent of Papers in Top Decile

by Subfield Publication Years: 1981-1984 Combined Journal Set: 1981 NIH-MEDLINE



7. Comparing the Performance of Grant Recipients with those whose Proposals were Rejected by the Agency

James G. Corrigan and Francis Narin, 1990

Objective

The citation frequency of three sets of papers was compared: papers produced by agency grant recipients, papers published by those whose proposals the agency rejected, and control papers. The question asked was whether the work funded by the agency was more highly cited than the work it did not fund.

Methodology

The agency supplied a list of proposals, principal investigator names and bibliographies associated with the proposals. CHI prepared author lists using these resources (due to varying formats and incomplete information on the bibliographies, they were used to glean author and journal names and not as a source of paper references). CHI searched for articles, notes and reviews published by these authors - the usual problems in compiling personal bibliographies applied. CHI looked up each paper in the library to obtain funding acknowledgements. Only papers acknowledging agency support were retained. For each of these a control paper was selected from the same journal, of similar length, without agency funding and concerning a similar topic (if the journal were multidisciplinary). Citation data were obtained.

Results

Five agency divisions and three publication years were examined, for a total of 15 division-publication year combinations. In 12 of the 15 cases, the median citation counts of papers published by grant holders were greater than or equal to both the median citation counts of those declined funding and the median citation counts of control papers. Divisions differed in that some had a "perfect" record - i.e. grantees' median citation counts exceeding declinees' and controls in all three years. One division's grantees obtained lower median citation counts than those declined funding. This division funded work in an area dominated by another granting agency.

Comment

This study was relatively bold in establishing comparable scientific work against which to compare agency funded research. The examination of work that was declined funding was also unusual and, as can be seen, provided valuable information.

(CH296)

8. Evaluating the Impact of a Technology Transfer Program on Scientists' Publishing

Francis Narin and Kimberly Stevens Hamilton, April 1994

Objective

A CRADA, or Cooperative Research and Development Agreement, is a written agreement between a company and a government laboratory to work together on a project. The CRADA is a vehicle devised by Congress to speed the commercialization of research and technology developed in Federal research. Underlying this study were questions about the quality of scientists who enter CRADA's and whether the CRADA compromises their scientific output.

Methodology

The study compared the publishing records of NIH scientists with and without CRADA's. Also publishing rates pre- and post-CRADA were analyzed to look for evidence of a drop in output. For each scientist with a CRADA, the client chose a scientist without a CRADA from the same institution and rank, or as near as possible. These matched scientists formed the control group. CHI compiled publication records for the two groups of 116 scientists using CD-ROM versions of Medline. This process was lengthy and problematic. Homonyms are a problem that is compounded by the listing of last names and initials only in databases. In addition, Medline lists only one address per paper, so essentially no address information was available to resolve questions. The subject of the papers was used to decide whether the paper belonged to the scientist in question. Tens of thousands of such judgments were needed to compile the bibliographies for the study and the attributions could not be perfect.

Results

The study found that scientists who enter CRADA's publish twice as many papers as the control scientists. CRADA scientists co-authored 8 to 12 papers per year while control scientists co-authored 4 to 6 papers per year. There was also some evidence, though not conclusive of a small downturn in publication rate after a scientist receives a CRADA. Taken as a whole, the group of control scientists showed no hint of a downturn in publication. The CRADA scientists show some evidence of a downturn though it is not significant statistically (2-tail significance of 0.2 to 0.25).

Comment

Given the difficulty of compiling individual bibliographies illustrated here, one would need a very strong result to be sure of the conclusion - such as the result that CRADA scientists are highly productive. The difficulties of tracking post-CRADA publication rates was compounded by the timing of the study. In the best case, only 3 years of post-CRADA publications were available even though the study was conducted seven years after the first CRADA (because multiple years of program operation were examined, publication lags must be accounted for, and current year publications cannot be used because of delays in journal publication). This illustrates a tension endemic to evaluation - that decision makers would like quick feedback, but research takes many years to have an impact on its field, and decades to have an impact on technology. The lags in publication are illustrative of the time needed for research to have an impact, but are often seen as a weakness in bibliometric method.

(CH357)

9. Benchmarking a laboratory's published output against world class competitors

Anthony F. Breitzman, September 1995

Objective

This study evaluated the laboratory's papers published between 1985 and 1994 using a variety of performance indicators to determine the laboratory's strengths and weaknesses and to compare the laboratory's performance to other similar organizations.

Methodology

A bibliography of papers listing the all variants of a laboratory's address was compiled from the *Science Citation Index*. A representative of the laboratory then helped to determine which papers were indeed from the laboratory and to classify the papers by laboratory section. Non-refereed papers (meeting abstracts, proceedings papers, chronologies etc.) were removed.

The following indicators were computed and analyzed by laboratory section, by field and subfield and in two time periods (1985-89 & 1990-94): number of papers, average research level (basic to applied), relative influence, relative citation index, number of highly cited papers.

Results

- The laboratory steadily increased publication output 67% between the two five year periods examined. Nearly all sections increased publication, but one particularly impressive section achieved a 246% increase.

- 5 of 10 sections account for 95% of papers.

- The laboratory ranks first among seven world class government labs in patents per million dollars, but is sixth in papers per million dollars. The laboratory equates to AT&T both in patents and papers per million dollars. The laboratory has the smallest R&D budget among the seven government labs examined.

- As a group the labs papers look somewhat mediocre on citation and other measures. However, the lab is doing world class research in mechanical engineering, solid state physics, optics and materials science. These fields relate to the labs core areas. The low overall figures are due to papers on the periphery of the lab's core specialties. It is common for labs to publish in peripheral areas and for these papers to be less highly cited.

- Shifts in emphasis are visible. Over the past 10 years the lab became stronger in physics and engineering and weaker in chemistry. Within physics there was an increase in the amount and impact of research in solid state physics and a decline in the impact of chemical physics

- Each of the sections with large numbers of papers have areas of strength and weakness and these were identified. Figure IV-5 displays the sections and their highly cited papers.

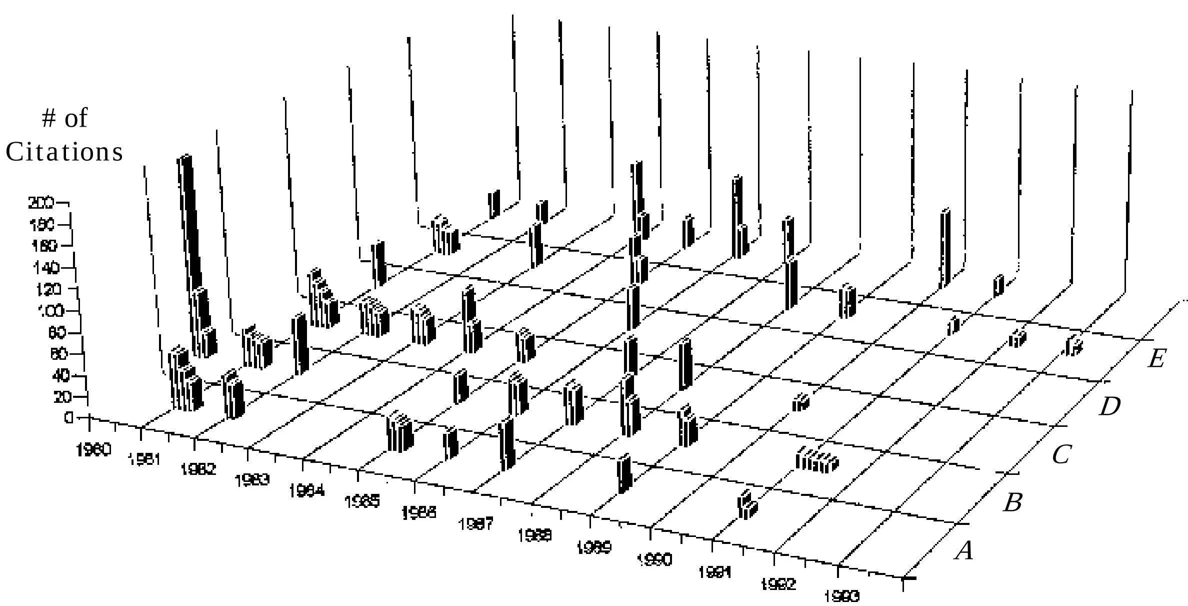
Comments

Publishing papers is not the primary mission of this laboratory, nor was this study intended as a complete evaluation of the laboratory's activities. Indeed, CHI also conducted an analysis of the laboratory's patent portfolio. Nevertheless, the study, which was conducted as an internal evaluation prior to reorganization, provides quantitative information on the laboratory's publishing output, identifying trends, areas of strength and weakness and making comparison's with other laboratories. Such information provides unique insights even though it may not be a comprehensive evaluation solution.

(CH404)

Figure IV-5.

Highly Cited Papers



*Papers with 25+ citations 1980-90 or 10+ citations 1991-94
**There are no papers with 10+ citations after 1993.
Note: Vehicle Prop., Vehicle Struct., Adv. Comp. and Survivability Directorates have no highly cited papers.

10. Evaluating the Impact of a Program to Enhance Research Capability in Traditionally Minority Institutions (RCMI)

Patrick Thomas, 1999

Objective

The RCMI program was set up to enhance the biomedical and behavioral research capabilities of academic institutions in which a large proportion of students come from minority groups. RCMI funding was first awarded to institutions in 1985. The purpose of the bibliometric evaluation was to establish whether the RCMI funding had led to an increase in the quality and quantity of research publications produced by these minority institutions.

Methodology

The evaluation was based on analysis of the publication records of institutions that had received RCMI funding for at least ten years. Two time periods were selected for the analysis – the period immediately before RCMI funding was awarded (1981-84) and a period after institutions had received funding for a number of years (1993-97).

A number of characteristics of institutions' publications were analyzed. These include number of publications; quality of journals in which papers are published; citation impact of papers; and percentage of papers co-authored with other institutions.

Results

The institutions were separated into three groups, depending on the extent of their research experience prior to the receipt of RCMI funding. The results showed that

- the least experienced institutions had not increased the quantity or impact of their publications. However, they had increased their level of co-authorship, suggesting that RCMI funding had allowed them to extend their links with the scientific community in general.

- RCMI funding had the greatest impact on institutions with a modest level of previous research experience. These institutions wrote more papers, published these papers in higher quality journals, and received more citations from later research.

- the publication records of institutions with the highest level of prior research experience did not change markedly after RCMI funding was awarded. This may be because RCMI funding represented a smaller percentage of their overall research budget.

Comments

In establishing that prior research experience may affect the impact of research funding, this bibliometric study provided information useful to managers designing future funding programs. In future, to achieve maximum impact on research output, it may be best to target this type of money on institutions with modest levels of research experience. The least experienced institutions may need another type of program.

The bibliometric evaluation formed part of a larger study into the impact of the RCMI program. The larger study included analysis of RCMI-funded institutions' success in competing for grants, their number of research staff and research students, and their development of infrastructures to support research.

(CH489)

11. Evaluating the Output of an Interdisciplinary Institute that Combines Science and Social Scientific Research

Diana Hicks, January 2000

Objective

To assess the international scientific impact of a multidisciplinary, social science oriented, foreign research institute receiving U.S. government support.

Methodology

The challenge in this study was to construct the bibliography of an institute where visiting scholars make large contributions. Because scientists may write up their work after leaving the institute, the institute's address might not appear on the paper. Therefore, three approaches to finding papers were combined: a survey of U.S. visiting scholars, the institute's own database of papers, a search for the institute's address on papers. The three sources had to be reconciled and cleaned to create an analytical database from which citation analysis could be conducted. Books and reports form an important component of the institute's output. Citations to these works were found and reported, though because no benchmarks were available, the information remained descriptive as opposed to analytical.

Once the bibliography was constructed, standard field classifications and citation norming procedures were used.

Results

The bibliography confirmed the institute's highly unusual research profile, with 59% of its papers in scientific fields and 41% in social scientific fields, a possibly uniquely balanced combination. Within science, its research profile is also unusual, combining earth science (17%), mathematics (8%) and engineering (11%) - these are smaller and less popular fields of research. This tells us that the institute pursues true interdisciplinary research, realizing an ideal often called for but infrequently practiced.

In a benchmarking of the institute's citation performance against field averages, the institute was found to be making an impact at the international level, with some outstanding areas and some very highly cited papers - see Figure IV-6.

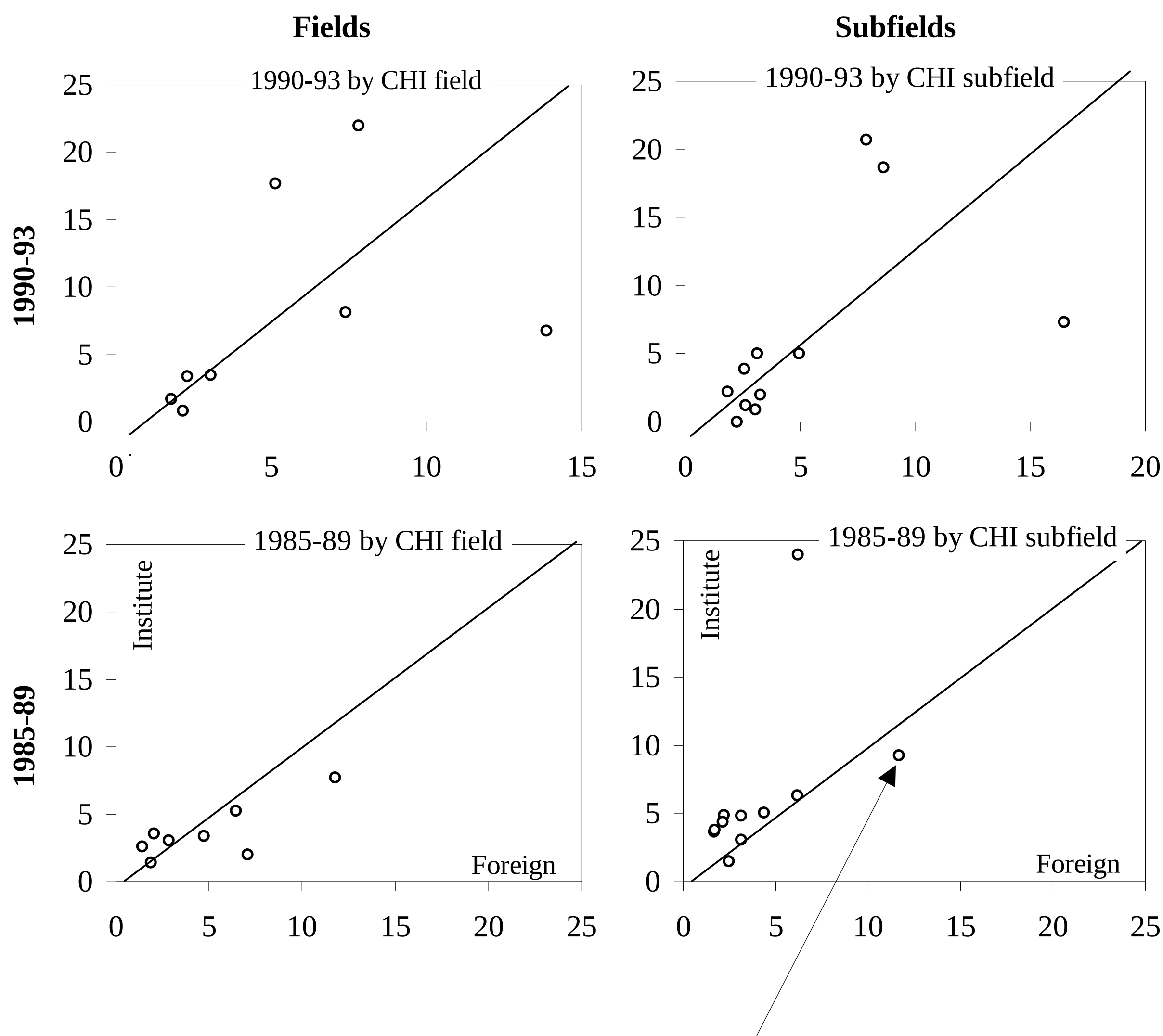
Comment

The reputations of non-mainstream institutes can often suffer unfairly. In its work, this foreign institute combines smaller scientific fields with social science fields in an unusual combination, though this result would have been stronger if the distribution of institute papers were compared against those of other non-university research institutes. Because quantitative indicators can be normalized for field and for foreign/domestic differences, a fair analysis could be produced that assessed the institute's contributions to each field.

(CH491.1)

Figure IV-6.

Citations per Paper – institute versus foreign norms



In each of the 4 graphs, the low field represents the Institute's papers in *Science* and *Nature*. The institute contribute mathematical modelling and environmental science to these journals whose citation norms are dominated by the much more highly cited fields of biomedicine. The diagonal is roughly the line on which the Institute's and foreign norms are equal.

12. Can Bibliometrics be Extended to the Web to Evaluate Outcomes?

Diana Hicks, January 2000

Objective

To see whether an institute's impact on the Web could be assessed

Methodology

The simple technique of searching on Alta Vista for mentions of the institute turned up about 6,000 hits. A closer look at some of these suggested that most were quite substantive, reflecting the use of the institute's work, or collaboration between the institute and a range of scholarly and policy organizations.

A second technique was examined in which referencing patterns on the Web were used to identify authoritative Web sites, i.e. those receiving many references from cites that were themselves authoritative. The algorithm used has a precursor in CHI's influence methodology developed by Pinski and Narin in the mid-1970's. The technique begins with pages found in a standard search engine search and augments them with pages that link to and from these pages. Several attempts were made and it was discovered that the most favorable results were obtained when precisely targeted search terms were available (for example "transboundary air pollution" rather than "environmental problem"), and the terms were related to pages on the institute's Web site that offered resources - such as software or data - as opposed to brochure type pages.

Results

The analysis of authoritativeness of Web sites suggested two things, first that the institute's resources were used in the policy sphere and second that the institute was almost unique. Institute resources are explicitly referenced on the Web, hence we know that they are influential. The institute's resources were used in the international policy sphere, as evidenced by the co-referencing of the institute's Web pages with those from international policy-making bodies. That the institute occupies a unique position in the world was deduced from the observation that there are few research organizations on the lists of Web authorities in areas of international environmental policy making. These authority lists contain instead government agencies and NGO's.

Comment

It is fortuitous for this type of institute that these techniques are becoming available, for they seem to confirm the institute's policy influence, something that has been impossible to validate quantitatively before. However, the lack of normalizations make it hard to know whether the results that were obtained are evidence of strong, normal or weak performance. The study also showed that the results obtained were extremely sensitive to the search phrase used. Furthermore, it was possible to influence the results by structuring a Web site with many pages that referenced each other. Finally, it has become well known in the cyberbibliometric community that search engine coverage is not stable day-to-day and the search features available change often. This is not conducive to high quality evaluation.

(CH491.2)

C.

Patent-to-Patent Bibliometrics:

Evaluating Research Outcomes

1. Evaluating the Impact of a Technology Transfer Program on a Scientists' Patenting

Francis Narin and Kimberly Stevens Hamilton, February 1994

Objective

This study was a bibliometric assessment of the impact of Cooperative Research and Development Agreements (CRADA) on the patenting behavior of National Institutes of Health (NIH) scientists. CRADA's were designed to facilitate the transfer of technology from national laboratories to industry through the mechanism of collaborative research.

Methodology

The patenting records of scientists who had received a CRADA were analyzed before and after the award was made, to see if the CRADA had any impact. In addition, each CRADA scientist was matched with a non-CRADA NIH scientist, as similar as possible in research institution and rank. This control group made it possible to analyze whether scientists involved with CRADA's were different in their patenting from scientists not involved in these agreements.

The patent data for the scientists included in the study were acquired using a straightforward, but somewhat time-consuming, approach. The name of each scientist was entered as a search term in the U.S. Patent Office records. The patents matching the name were then evaluated to see which of them were invented by the scientist in question. This was achieved by checking the assignee of the patent, and its subject matter.

Results

Throughout the period being studied, the CRADA scientists were granted an average of five times as many patents as the control group containing non-CRADA scientists. Their patents were also more highly cited than those of the control group. In terms of cycle time (innovation speed) and links to science, the patents of the CRADA and non-CRADA scientists were similar.

The CRADA scientists increased their rate of patenting after receiving their agreements. Over the same time period, the control group also produced more patents. The study therefore concludes that CRADA itself did not have a substantial impact upon the patenting behavior of scientists. However, scientists who receive CRADA's are likely to have a different orientation toward patentable biomedical research both before and after receipt of a CRADA.

Comments

This study shows how bibliometric analysis may be used to evaluate the impact of programs designed to enhance technology transfer. One point of interest about the study is the approach it uses to identify the patents of individual scientists. This approach (described in the Methodology section) is not perfect, and may run into problems where there are two scientists with the same name working in the same research area or at the same institution. This highlights the problems associated with using patent or publication data at the level of the individual scientist.

(CH354)

2. Co-inventor Clusters Applied to Intellectual Property Management

John Perko and Francis Narin, December 1994

Objective

This project was carried out for the Naval Research Laboratory (NRL). Its purpose was to improve NRL's management of its intellectual property, notably its patents. This involves maximizing licensing revenues from the patents, recognizing the contributions of different researchers within the laboratory, and protecting NRL's rights to its inventions.

Methodology

NRL's patents were split into groups based on their inventors using CHI's brain mapping software. The idea behind brain mapping is to identify clusters of inventors who are linked through their co-invention of patents. The groups identified by the brain maps (Figure IV-7) tend to be working on a specific problem within a particular technology. The brain maps also identify the inventors who appear to be central to the research efforts in a particular area, reflected in the large number of patents on which they are listed as an inventor.

Having split the NRL patent portfolio into co-inventor clusters, these clusters were then analyzed to identify the most highly cited patents. These patents were regarded as being likely candidates for licensing, based on the assumption that highly cited patents tend to contain important technological information.

Results

The brain maps highlighted the work of a number of inventors with highly cited patents. These inventors are working on technologies that are being used by number of other organizations. This has two implications – firstly that these patents are a good candidate for licensing, and secondly that these inventors should be supported in their future research efforts.

The analysis also revealed that NRL's patents in general were cited more often than average. This suggested that NRL was producing technology that was having a strong impact on later developments. A number of NRL patents also cited a large number of scientific papers, showing their use of the latest scientific developments.

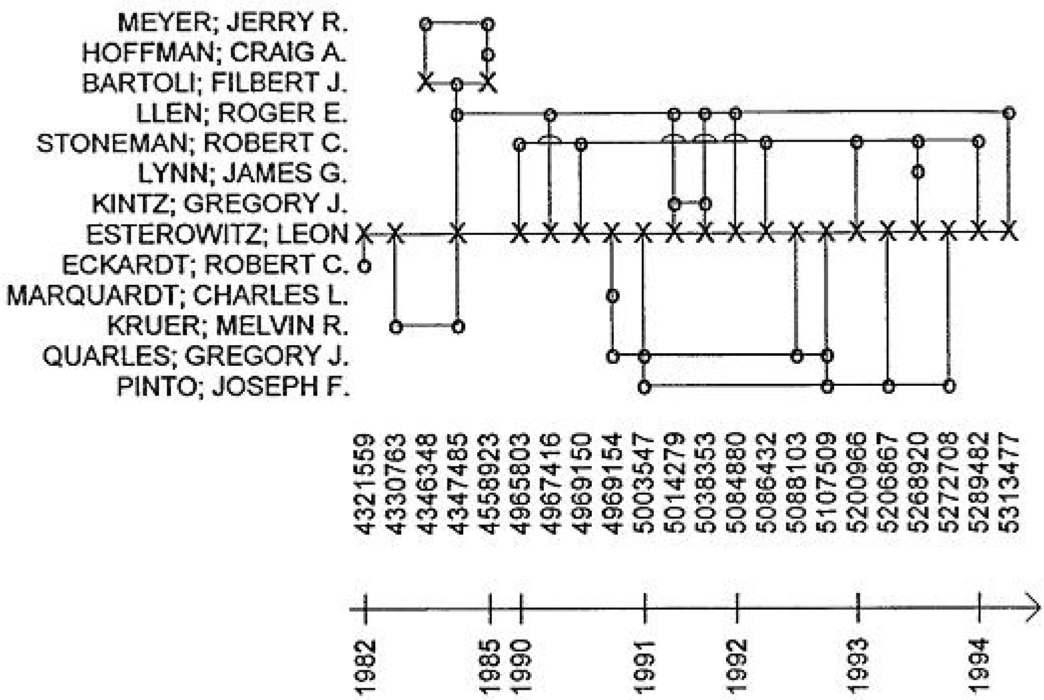
Comments

Brain mapping is a useful technique, both for identifying key inventors within organizations, and for splitting patent portfolios into different technology groupings. It can also be used to determine whether particular inventors are still with a particular organization, or have moved elsewhere. This can be important information, given that the patenting of most organizations often depends on a small percentage of their inventors.

(CH384)

Figure IV-7.

Co-Inventor Cluster



3. Assessment of a Laboratory Patent Portfolio

Anthony Breitzman, Margaret Cheney and John Perko, December 1994

Objective

The project was designed to compare the patent portfolio of the Army Research Laboratory (ARL) with a variety of leading worldwide laboratories. In particular, its purpose was to identify strengths and weaknesses of the ARL portfolio, and to compare these strengths and weaknesses with those of laboratories working in similar areas.

Methodology

ARL patents were identified with help from ARL itself. This is because all U.S. Army patents are assigned to USA Secretary of the Army, rather than to the individual parts of the Army. A group of 14 government, university, commercial and foreign laboratories were selected as the comparison group. These included the Naval Research Laboratory (NRL), AT&T Bell Laboratories, Stanford University and the Toyota Research Center. In order to identify the patents of these 14 laboratories, a combination of assignee names, inventor names and addresses were used.

Patents within a number of SIC Product Groups were selected for analysis. These groups included Electronic Components & Accessories, Professional & Scientific Instruments, and Non-Electrical Machinery. These patents were analyzed using CHI's patent indicators, in order to determine their impact, speed of innovation and links to science (Figure IV-8).

Results

Among the government laboratories, ARL and NRL had the largest number of patents. However, according to CHI's patent indicators, NRL's patents were strongest among the government laboratories, while those of ARL were of average strength compared to this group. ARL had the weakest links to scientific research, suggesting that its patents were not building on the latest scientific developments.

All of the government laboratories tended to have weaker patent indicators than the commercial laboratories. This may be because of the more specialized nature of the patents produced by the government laboratories.

The one positive finding for ARL was that it produced more patents per active inventor than any of the comparison laboratories. This suggests that ARL's inventors were more productive than their counterparts in both public and private laboratories.

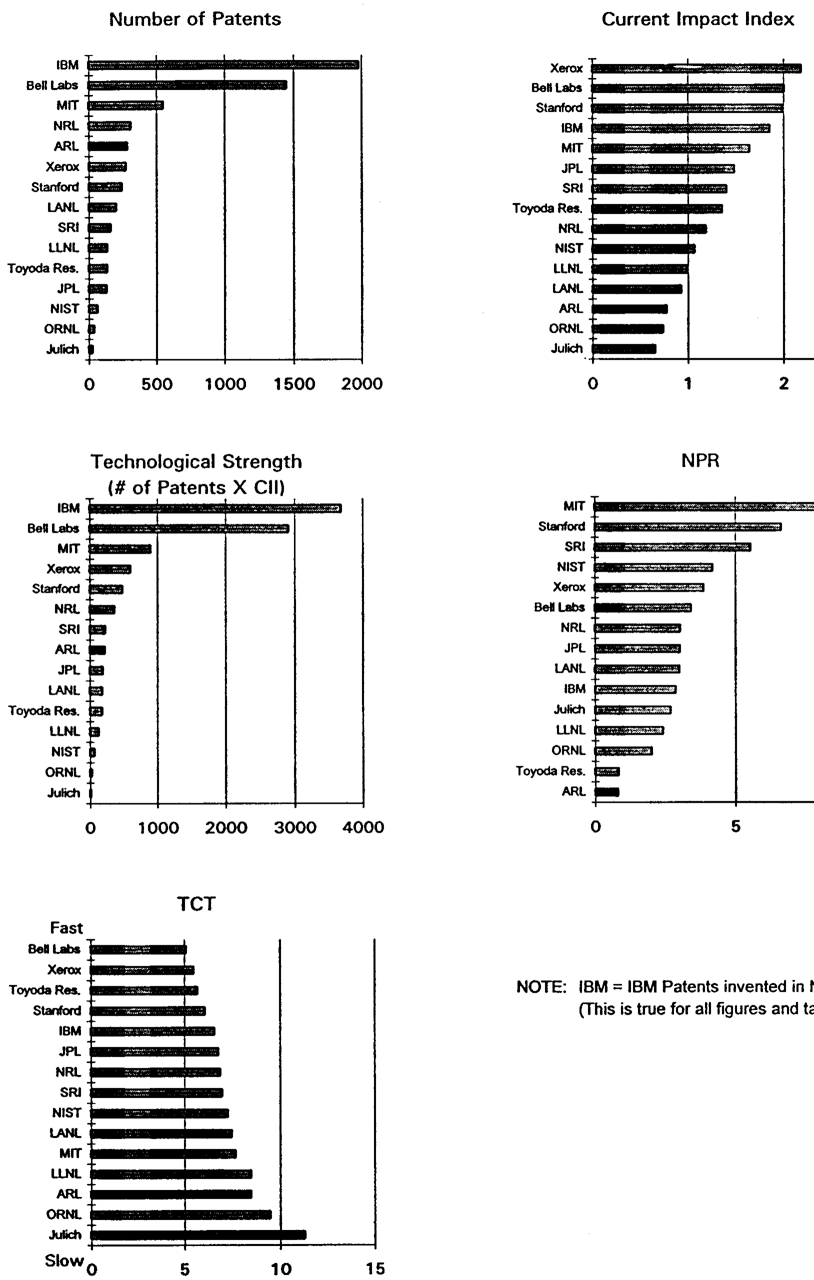
Comments

This study shows how patent indicators can be used to compare the entire patent portfolios of a given set of organizations. They provide quantitative insights into the relative strengths of different organizations, without requiring extensive analysis of individual patents. The most time-consuming aspect of a study of this type is the construction of accurate patent lists for each organization.

(CH387)

Figure IV-8.

Patent Activity 1990-94



NOTE: IBM = IBM Patents invented in New York.
(This is true for all figures and tables)

4. Assessment and Augmentation of Technology Transfer through Citation Analysis of a Patent Portfolio

John Perko, June 1995

Objective

This project was an evaluation of the patent portfolio of the Naval Command Center & Surveillance Center (NRAD) of the U.S. Navy. Its purpose was to provide an overview of the quantity and quality of NRAD's patented technology, and to identify patents that may offer technology transfer opportunities.

Methodology

NRAD supplied a list of their patents to CHI for analysis. CHI then categorized the patents based on their Product Group, a set of codes related to the Standard Industrial Classification (SIC) codes. Examples of Product Groups are Electronic Components; Electronic Transmission & Distribution; Professional & Scientific Instruments; and Office/Computing/Accounting Machinery. The U.S. Patent Office assigns Product Group codes to individual patents.

The citation patterns associated with the patents were then analyzed. This analysis included comparison of the quality and quantity of NRAD's patents in each category, and identification of the patents that were linked closely to scientific research. It also included a discussion of the areas of NRAD's technology that were being cited frequently by patents from other organizations. These technologies may be candidates for technology transfer.

Results

NRAD obtained patents at a relatively steady, low rate throughout the 1980s, before increasing their rate of patenting from 1990 onwards (Figure IV-9). The biggest increase in NRAD patenting came in Optics, Fiber Optics and Electrical Equipment.

NRAD's early patents were highly cited, relative to other patents from the same technologies and years. However, their later patents were less highly cited. This suggests that while NRAD started to produce more patents after 1990, these patents did not attract as much interest from other organizations.

Over half of the citations to NRAD's patents were from private U.S. companies, and a further third were from foreign sources, mainly private companies. This suggested that there were significant opportunities for NRAD in technology transfer, licensing and alliances with private companies. A list of the companies citing NRAD's technology on a regular basis provided possible targets for such an initiative.

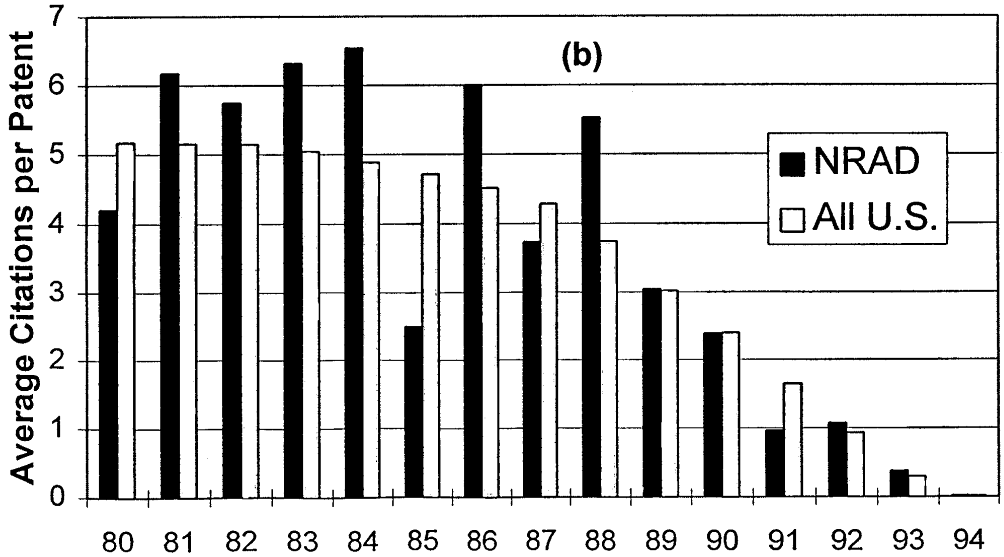
Comments

This study shows how citation techniques can be used to aid the management of patent portfolios. CHI undertook similar analyses of the patent portfolios of other parts of the U.S. Navy.

(CH398)

Figure IV-9.

Overall NRAD Patent History 1980-94



5. Evaluating the Contributions Made by a Federally Maintained Database to Research and Technology

Michael Albert, December 1996

Objective

This study evaluated the contribution of a federally maintained database to research developments in a particular field. The database analyzed was the standard reference database Thermodynamic Properties of Refrigerants and Refrigerant Mixtures (REFPROP) maintained by the National Institute of Standards and Technology (NIST). The research field examined was CFC (chlorofluorocarbon) replacement.

Methodology

The study was designed to evaluate the impact of the REFPROP database on publishing and patenting in CFC-replacement. It was therefore necessary to define that research field. This was achieved using a technology filter comprising keywords and, in the case of patents, International Patent Classifications.

An additional stage was added to the process of identifying relevant patents. This involved examining all of the patents that cited to the patents identified by the initial filter, and determining which of them should be included. These additional patents were building on CFC-replacement technology, and may represent the next generation of refrigeration technologies that do not mention CFC's explicitly.

The patents and papers were then split into two groups, according to whether or not the organizations producing them subscribed to the REFPROP database. This grouping allowed for a comparative analysis of the impact of the REFPROP database.

Results

The study showed that REFPROP subscribers held just over half of the U.S. patents, and just under half of the scientific and technical papers, in CFC-replacement technology. The influence of REFPROP was particularly strong among the large patenting and publishing organizations. Out of the top ten patenting organizations in this field, seven were subscribers to REFPROP, including four out of the top five. Similarly, five out of the top ten publishing organizations in this field were subscribers, including all three commercial organizations in the top ten.

REFPROP subscribers were also responsible for almost two-thirds of the highly cited U.S. patents in CFC-replacement (Figure IV-10). This is a higher percentage than expected, given that REFPROP subscribers account for just over half of the U.S. patents in this area. It suggests that REFPROP subscribers tend to produce research whose impact is greater than average.

Comments

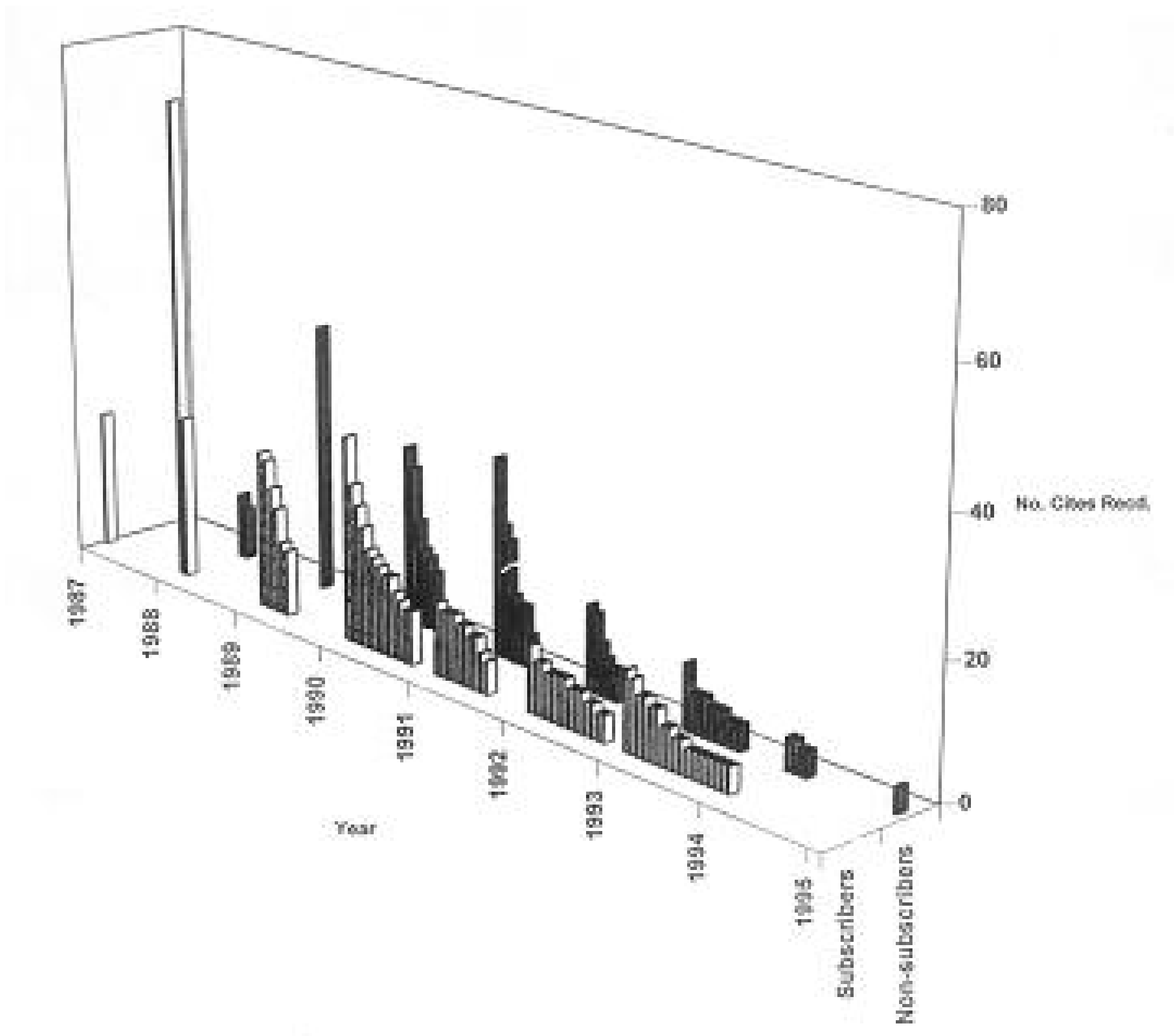
This study extended bibliometrics in an interesting way to assess the impact a federally maintained database had on research and technological development. The impact of resources like databases is usually assumed not amenable to bibliometric analysis. However, this bibliometric study showed how a federally maintained database can be an important source of support for research and development.

(CH428)

Figure IV-10.

"Stick Diagram" Comparing Incidence of Subscriber and

Non-Subscriber Top-Decile CFC Replacement Patents



6. Patent Citation Mapping for Government Patent Portfolio Management

Anthony Breitzman, Margaret Cheney and Francis Narin. July 1998

Objective

This study aimed to provide information that could help improve the management of the National Institutes of Health (NIH) patent portfolio. In particular, the study focused on identifying patents and clusters of patents owned by NIH that were highly cited. These patents may be candidates for licensing.

Methodology

NIH provided CHI with the list of patents in the NIH patent portfolio. With help from NIH scientists, CHI then grouped these patents into seven categories using a combination of U.S. Patent Office Classification (POC), Standard Industrial Classification (SIC) Product Groups, and keywords. The seven categories were Genetic Engineering, Bio-Affecting Drugs, Other Drugs, Diagnostics, Instrumentation, Organic Chemistry, and Other.

The patent categories were compared in order to identify strengths and weaknesses within NIH's patent portfolio. At a more detailed level, individual highly cited patents were highlighted, along with the organizations that cite them. The patents are potential candidates for licensing, and the citing organizations are potential licensees. Inventors responsible for large number of patents or highly cited patents were also identified.

Results

The main finding of this study was that NIH patents were comparable to those of leading biotechnology and pharmaceutical companies, in terms of citation impact and links to science. Also many of NIH's patents were cited frequently by corporate enterprises, suggesting that they may offer licensing opportunities.

Particularly highly cited were NIH's patents in Instrumentation. On average, they received approximately twice as many citations from industry as patents in most other categories.

The inventor analysis revealed that a small number of scientists were responsible for a relatively large percentage of NIH's patents (Figure IV-11). A similar concentration of productivity has been reported in many research organizations. This shows the importance of a small number of key inventors.

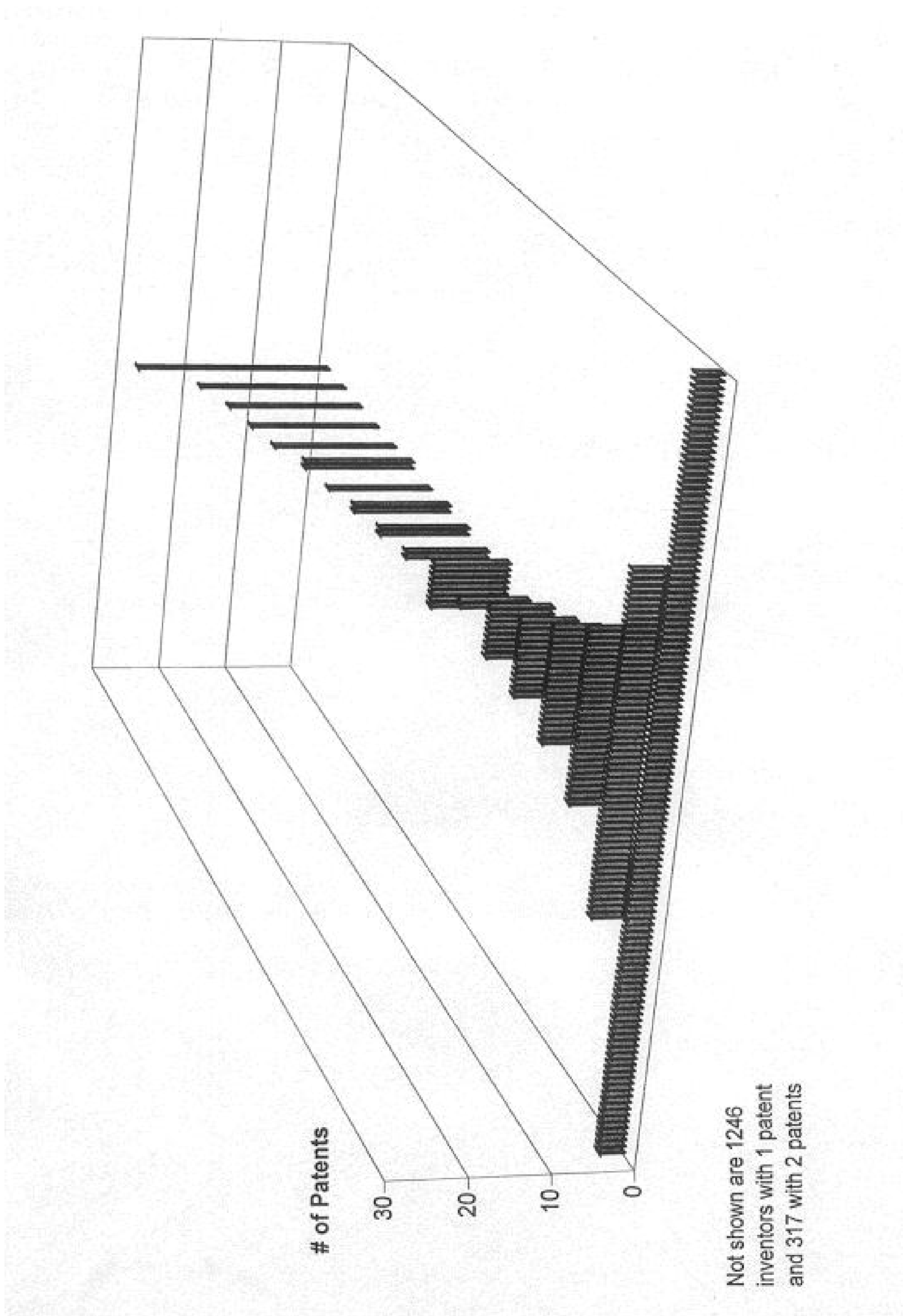
Comments

Managing a very large patent portfolio, such as that of NIH, can be a difficult task. This study reveals various ways in which a patent portfolio can be analyzed in order to identify technological strengths and licensing opportunities. These techniques may be used by federally funded bodies, such as NIH, to generate additional revenue.

(CH464)

Figure VI-11.

Number of NIH Inventors VS. Number of Patents



D.

Patent-to-Paper Bibliometrics:

Evaluating Outcomes Using Evidence of Research-Technology Linkage

1. In a Science-based Technology, a Nation's Presence Fades Away as Application Approaches

Francis Narin & Dominic Olivastro, August 1994

Related paper: J. Anderson, N. Williams, D. Seemungal, F. Narin and D. Olivastro. "Human Genetic Technology: Exploring the Links Between Science and Innovation," *Technology Analysis and Strategic Management*, 8, 2, 135-156, 1996.

Objective

This study characterized the patented technology and related research related to genetic engineering and human genome technology (HGT) and delineated the role of U.K. inventors and scientists therein.

Methodology

To construct the patent set for human genome technology, a filter was applied to the USPTO patent system. The filter combined patents in 18 IPC classes, more than 25 USPOC classes, listing 8 keywords in combination with gene or genetics or 6 other keywords. The resulting 1,500 patents were scanned by CHI and the client to remove patents obviously non-human in their nature or application, and to try and remove patents which were primarily chemical modifications of materials such as nucleotides and nucleosides and other chemical compounds patented as therapeutic agents but not informed by human molecular and cellular data. The final set comprised 1,105 patents.

An attempt was made to classify the final set of patents by both novelty and application. The novelty classes were: 1) nucleic acids and short or modified sequences of DNA/RNA; 2) Proteins and production processes for hormones/lymphokines/growth factors; 3) monoclonal antibodies and their production processes, and 4) viruses/vectors/vaccines. The application categories were: 1) assays; 2) anti-microbial; 3) cell growth and culture; 4) cancer-related; 5) other and 6) production techniques and not stated.

Data acquisition was completed by pulling the patents and papers cited by the HGT patents and the patents that referenced the HGT patents. The locations of the patents' inventors, the number of references on the patents, the location of paper authors and the funding acknowledgements on the papers (looked up in the library) provided the basic data used in the study.

Results

The UK presence became progressively weaker as one generation of patenting was succeeded by another. That is as the technology presumably approached application, the U.K. faded away - see Figure IV-12. In the first generation, that is material cited by HGT patents, the U.K. achieved a 6.4% share of papers and 3.2% share of patents. In the HGT patents themselves, the U.K. share slipped to 2.8%. In the patents citing HGT patents, the U.K. share was down to 2.5%.

The study also found that on average the HGT patents cited more papers (8) than patents (3). It also found that inventors heavily over-cited their own country's scientific papers, illustrating the strong in-country component to relations between science and technology.

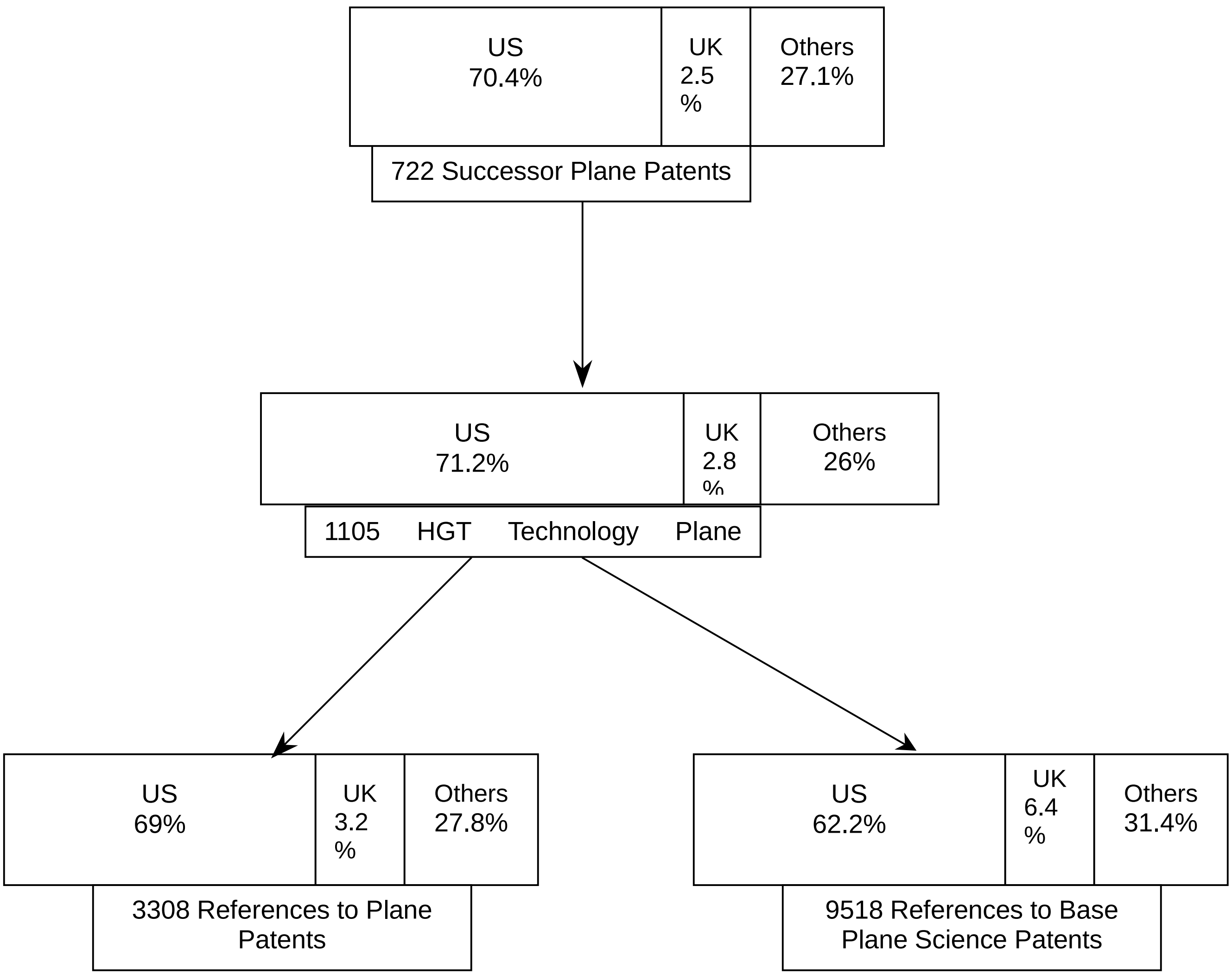
Comment

The first challenge in the project was to define the patent set representing human genome technology. Clustering and artificial intelligence methods can be used to solve this problem, however it is also straightforward to solve it with the application of real intelligence and knowledge. The reward for doing so is certainty that the resulting patent set contains all the patents needed and only those patents (high recall and precision).

(CH376)

Figure IV-12.

Overview of HGT 3-Plane Model



Note: Arrows indicate direction of Citation

Note: Additional 5000 Non-Patent References of various kinds

2. Benchmarking an Institute's Patenting against Similar Laboratories

Francis Narin and John Perko, August 1994

Objective

This study examined 15 leading laboratories in the U.S. Its focus was the relationship between the scientific research undertaken by these laboratories, and the patented technologies that build on this research. The purpose of the study was to show how scientific research had an impact on technological development.

Methodology

The front page of a U.S. patent contains citations to previously published items that limit its claims. These items represent the technology upon which the current patent builds. They are separated into two groups – references to existing patents, and references to non-patent literature. Most of the latter group is made up of scientific papers, and forms the basis for this study.

To identify the papers written by the 15 laboratories, this study used an existing science-technology linkage database developed by CHI. To construct this database, CHI took the 130,000 references to non-patent literature from all U.S. patents issued in 1987 and 1988, and directly matched them to the journals covered by the Science Citation Index. Approximately 30,000 of the references were matched using this approach, along with the institutional addresses of all the U.S. authors of these papers.

Results

Figure IV-13 shows that out of the 15 organizations studied, the papers written by the National Institutes of Health (NIH) received the highest number of citations from patents. Following NIH were IBM, the U.S. Department of Agriculture (USDA), Lincoln Laboratories and the Naval Research Laboratory.

The areas of dominance of the different laboratories were as one would expect. NIH was dominant in biomedicine and clinical medicine; USDA had the most highly cited papers in biology; and IBM had a large number of citations in Physical Science and Engineering.

There was a strong national component in the citation patterns. For nearly all the laboratories, 70-90% of the citations they received were from organizations based in the U.S. Since the foreign share of all U.S. patents was about 50%, this indicates that U.S. science is feeding U.S. technology more than overseas technology. This is consistent with other findings showing that inventors preferentially cite the science of their own country.

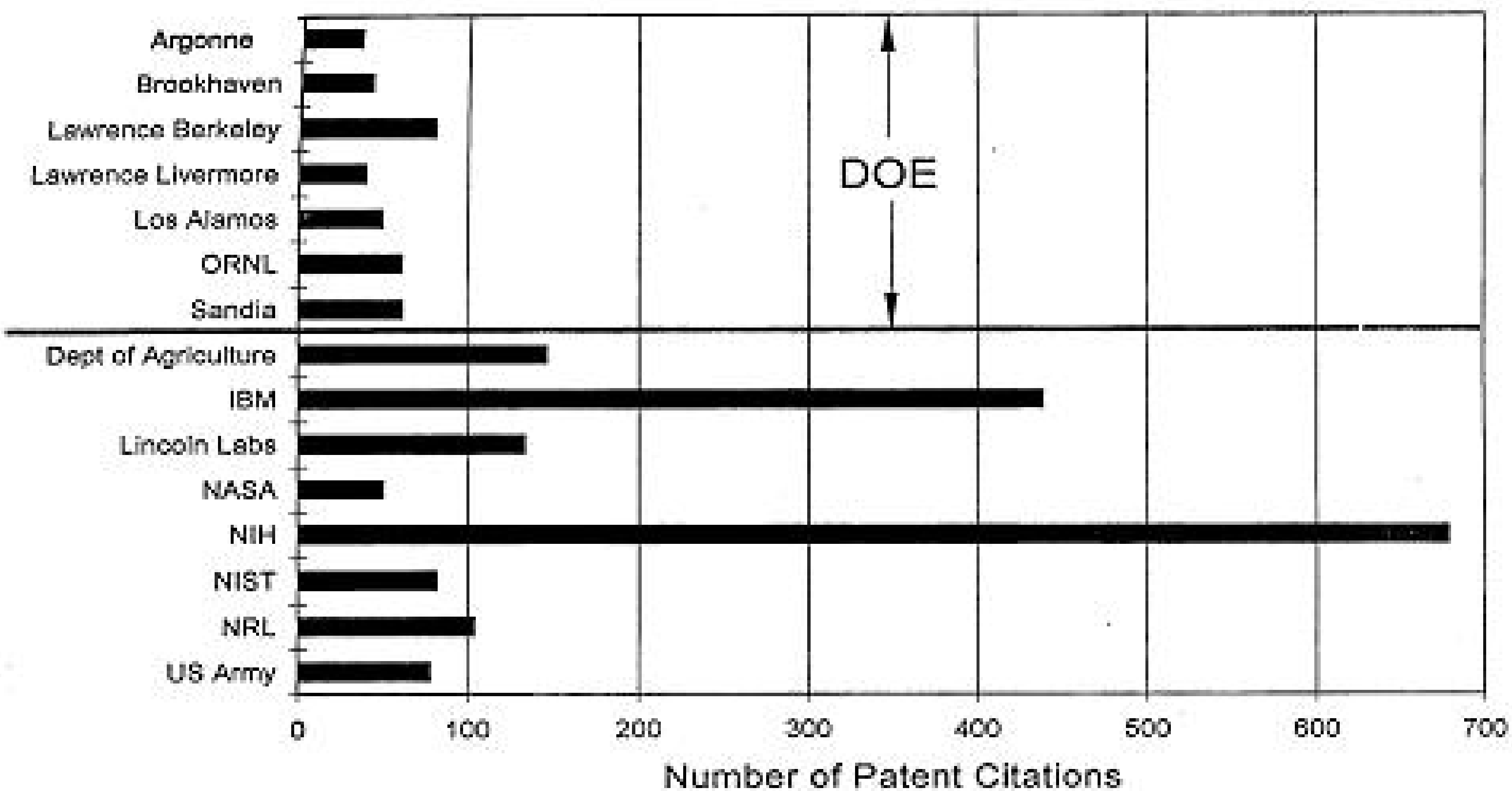
Comment

The citation linkage between scientific papers and patents provides a way of quantifying the impact of scientific research on technological developments. This study shows how this type of data may be used to show the importance of undertaking scientific research.

(CH372)

Figure IV-13.

1987-1988 Patent Citations to 1973-1986 Lab Papers



3. Examining Institute Papers Cited in Patents

Parallel studies in three reports: Francis Narin, Peter Kroll, and Anthony Breitzman all dated January 30, 1996.

Objective

Three institutes contracted for studies of the contribution made by their funded research to patented technology. Both the quality and quantity of the cited science and resulting citing technology were of interest.

Methodology

Taking a data set of all papers with at least one US author and cited by US patents issued in 1987-88 or 1993-94, those papers were examined in academic libraries for acknowledged support source and author institution, and papers affiliated by any of the three institutes—either through funding or institutional affiliation—was extracted for further analysis. Indicators and time trends for patents and papers were studied, as well as differences between intramural and extramural authorship.

Results

Between 1987-88 and 1993-94, the number of patents citing papers supported by the institutes increased by a factor of roughly 3, as did the number of individual papers cited patents (Figure IV-14). The total number of such citations increased even more. Top cited author institutions reflected the institutes' intramural or extramural focus, and the research most cited by patents. Top assignees—both universities and research institutions, and private corporations—show the companies or research organizations benefiting most from the Institute-related research. Those Institutes with strong intramural programs showed significant levels of patenting on their own.

Institute-related papers most highly cited by patents were identified to indicate the significance of particular research subjects or authors. Papers were grouped by subfields and level and influence characteristics, giving an indicator of research quality. Non-Institute support acknowledgments of other sources were tabulated to indicate joint funding sources.

Citing patents were also characterized by assignee sector and country, and the patents that were themselves most highly cited by future patents were identified as potentially significant technological advances that were traceable to the original Institute-supported work.

Comments

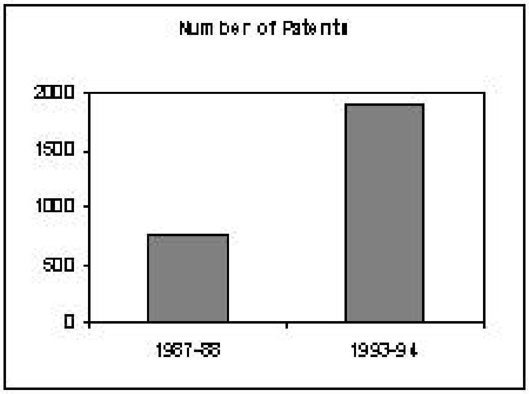
A major product of this work was creation of a relational patent-to-paper linkage database, useful for analyzing the multi-stage relationships of research funding through to technology from a variety of aspects.

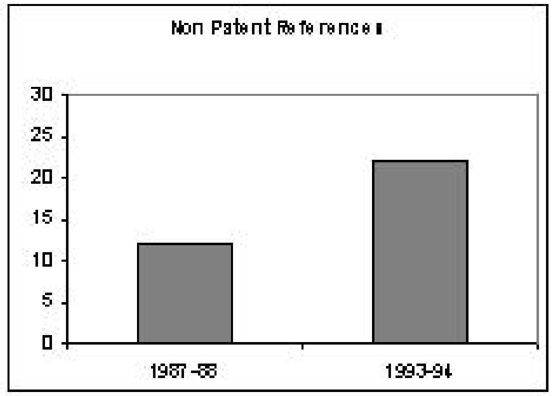
(CH412, CH413, CH414)

Figure IV-14.

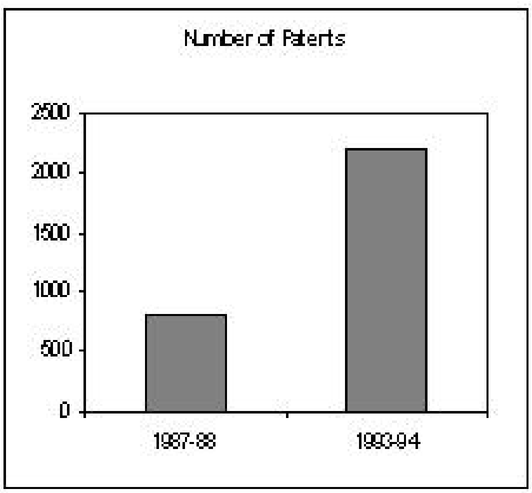
Indicators for 1987-88 & 1993-94 Patents Citing Funded/Authored Papers NIGMS

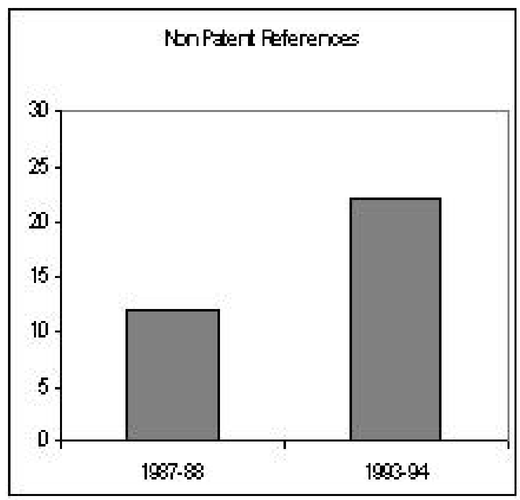
NIGMS



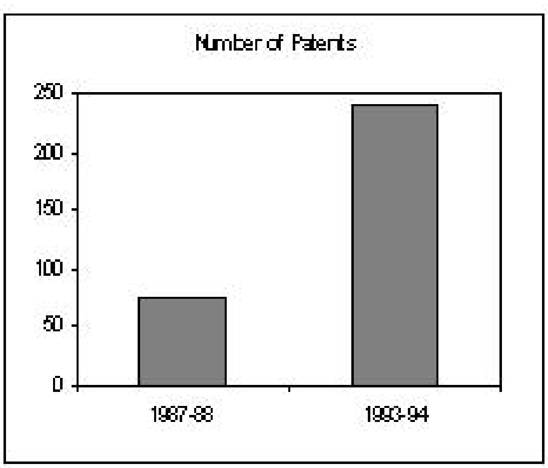


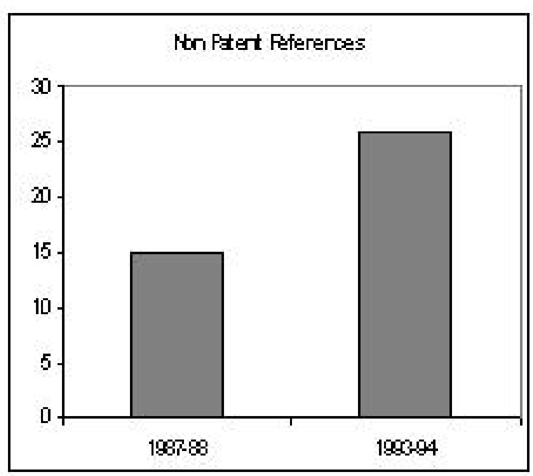
NCI





NIMH





4. Characterizing the Patented Technology that References Agency Science
John Perko, October 1996

Objective

This study examined the contribution made by Agricultural Research Service (ARS) papers to patented industrial technology.

Methodology

All papers that met two criteria were identified: 1) originating at ARS laboratories or funded by the U.S. Department of Agriculture (USDA) and, 2) cited in 1987-1988 & 1993-94 U.S. patents.

Results

The number of papers and citations increased three-fold between 1987/88 and 1993/94. About half of USDA supported papers were from ARS laboratories and half from universities. The ARS papers were more applied and agricultural in orientation and the university papers were more basic and biomedical. Patents citing these papers cited a large number of scientific papers, on average 10 or more. This compared to the U.S. average of 1 citation to a scientific papers per patent. Roughly half the U.S. patents were foreign owned at this time, but more than 90% of the patents citing ARS papers were invented in the United States, demonstrating that ARS science contributes to U.S. technology.

All papers published in 1988 and cited in patents issued through 1995 were examined. In the field of biology, the institution with the most papers cited in patents was ARS with three times more papers cited than any other single institution.

The top organisations whose patents referenced ARS papers were also identified - see Figure IV-15 The product groups in which patents most frequently reference ARS papers were also identified - see Figure IV-16.

Comment

CHI undertook a number of similar studies. Patent-to-paper bibliometrics makes a valuable contribution to the assessment of research outcomes (as opposed to outputs). The results are easy to understand, being counts, and do not rely on subjective answers to "what if" questions to assess effects. Counterbalancing this is the disadvantage that a dollar figure cannot be assigned to the research uses identified.

(CH425)

Figure IV-15.

Top Ten Patent Assignees Citing ARS Papers

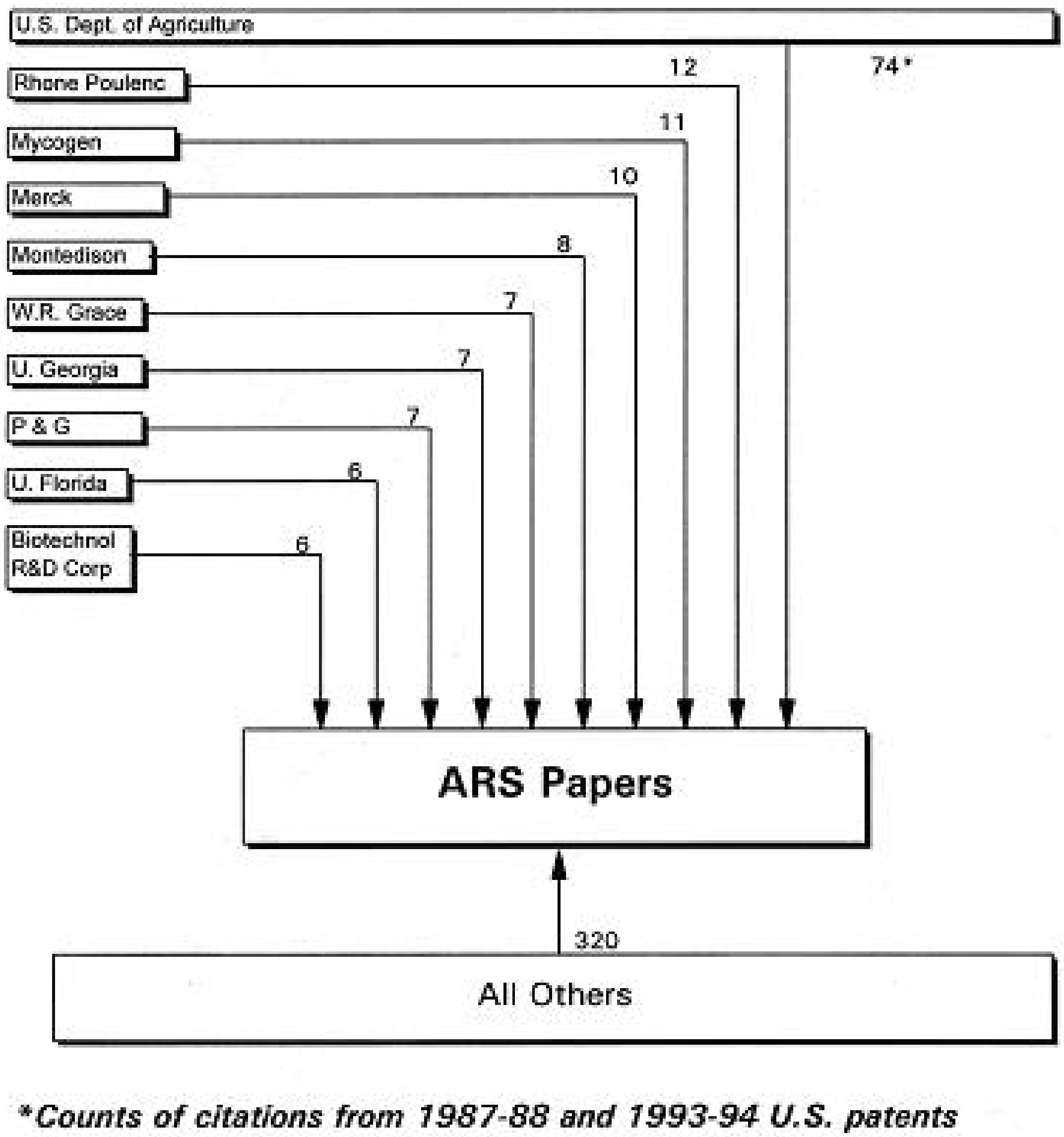
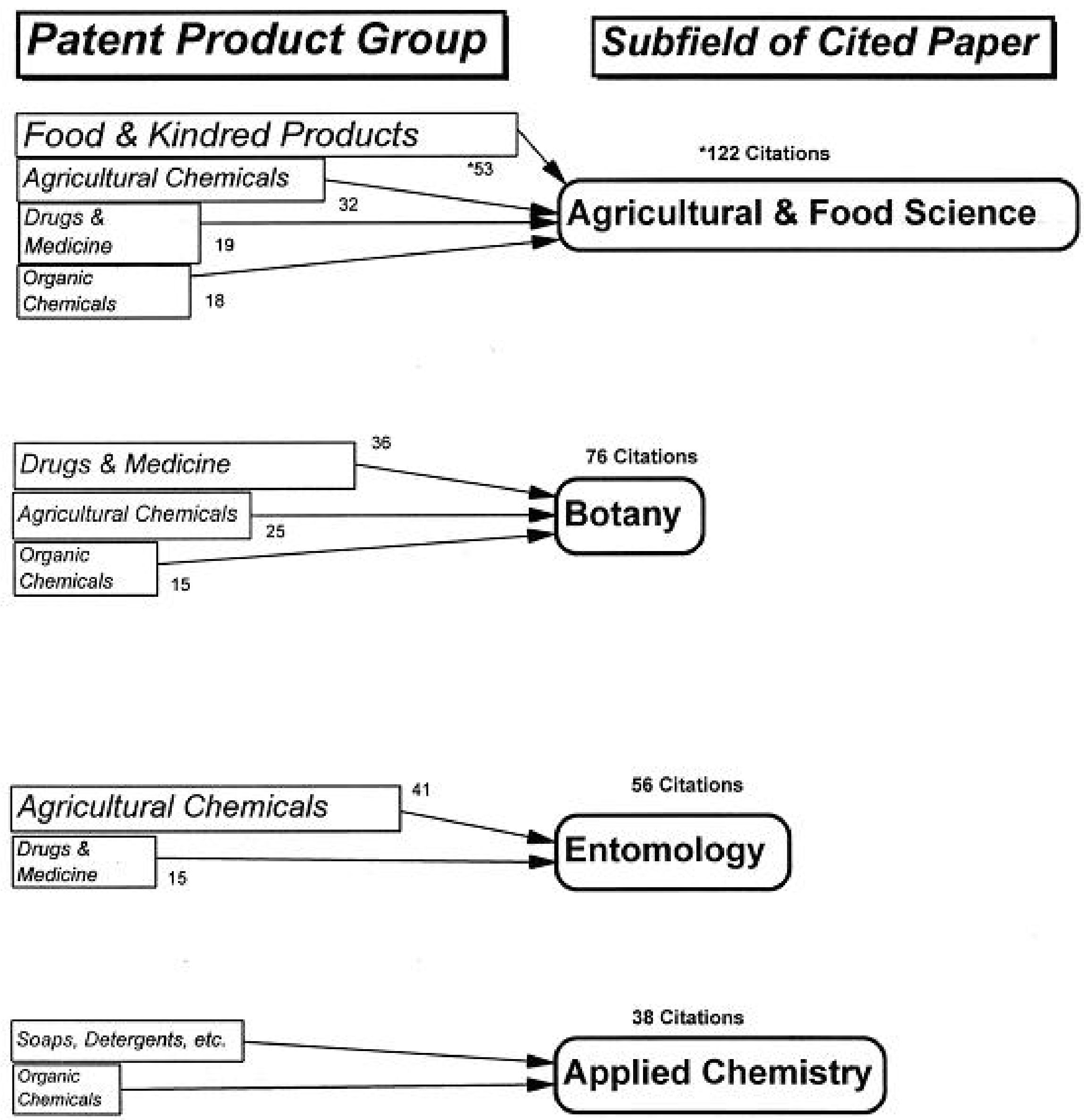


Figure IV-16.

Most Frequent Citation Connections

Between Patent Product Groups and ARS Publishing Fields



***Counts of citations from 1987-88 and 1993-94 U.S. patents**

5. Examining the importance of NEI-funded research to eyecare technology

Leon B. Ellwein, Peter Kroll, and Francis Narin, "Linkage Between Research Sponsorship and Patented Eye-Care Technology," Investigative Ophthalmology & Visual Science, vol. 37, no. 12, November 1996, 2495-2503.

Objective

The National Eye Institute wanted to assess the state of eyecare technology by studying patenting from 1975-96; specifically the determination of NEI-funded research on the patented technology.

Methodology

8,163 eyecare technology patents were identified from 1975-94 US patents using a definitional filter based on Patent Office Classifications and keywords. The science references cited by those patents were looked up in academic libraries, with author institutions and funding organizations recorded. The study assessed trends over time, institutional authorship, and other aspects of the patents, but the area of interest here is research funding in the cited papers.

Results

Table IV-4A tabulates number of patents, science references, cited papers, and support acknowledgments for each support source category (NEI, other NIH, and other broad categories). Of the 1,192 patents for which science references were found, 739 (62%) had a reference to at least one paper where some form of external support was acknowledged. Considering science references instead of patents as the unit of interest, of a total of 4,889 found science references, 2,588 (53%) were to papers with one or more acknowledged funding sources. As a single institution, the NEI plays a major role with 371 patents (31%) containing 1001 science references (20%) to 606 different papers (18%) with an acknowledgment of NEI support. For only those patents, references, and papers where some source of support was acknowledged, NEI is represented in 50%, 39%, and 34%, respectively. Of 6130 total acknowledgments, NEI is credited with 28%. When NEI is merged with the rest of NIH (i.e., the NIH is treated as a whole), the overall percentages reach 41% of patents with science references linked to NIH, 34% of science references linked to NIH, 33% of cited paper linked to NIH, and 51% of support acknowledgments linked to NIH. Taken as a whole, the US private nonprofit sector is comparable to the NEI as a major support source.

Table IV-4B shows the distribution of these patent references to science broken down by type of author institution and funding support source. NEI and NIH support a large fraction, 34%, of the entire research portfolio, but that support is particularly strong (59%) for research with the US academic sector.

Comments

The method of identifying links is relatively conservative, by starting with a restricted patent set. Research tool patents relevant to the area may not have been selected as being related to eyecare technology if patent classes or keywords did not fall within the strict classification filter definitions.

(CH460)

Table IV-4A. Patents with Science References, Papers, and Acknowledgments by Support Source Sector for Eye Technology Patents: 1975-1994

Support Source	***Number of Citing*** Patents with Science References	Number of Science References	Number of Cited Papers	Number of Support Acknowledgments in Cited Papers
National Eye Institute*	371	1001	606	1687
Other NIH*	246	790	607	1452
Other US government	250	407	300	501
US private nonprofit	389	892	625	1278
US foreign for-profit	230	324	225	376
Foreign government and nonprofit	268	542	412	798
Not fully identified	32	36	29	38
Any acknowledgment	739	2588	1808	6130
No acknowledgment	453	2301	1616	NA
Total	1192	4889	3424	NA

NA = not available.

*Combine NEI and Other NIH into a single category results in the following counts: patents = 491; references = 1674; papers = 1133; acknowledgments = 3139. For patents, references, and papers, these counts are less than the sum of the two rows in the table because of the avoidance of double counting when a single paper acknowledges both an NEI and a non-NEI NIH institute.

Table IV-4B. Science Reference Counts Disaggregated by Author Institution and Support Source for Eye Technology Patents: 1975-1994

	Author Institution								
Funding Support Source	US Academic	Foreign Academic	US for Profit	Foreign for Profit	Federal Government	Other US	Other Foreign	Total	No Information
NEI*	726.2	41.7	9.2	2.0	22.5	189.7	6.8	998	3
Other NIH*	565.0	27.3	25.7	5.0	38.7	118.0	9.3	789	1
Other US government	289.1	9.3	15.2	3.0	29.1	55.2	2.1	403	4
US nonprofit	600.7	41.7	17.0	1.0	34.3	186.0	10.3	891	1
US + foreign profit	182.2	40.2	28.2	1.0	6.8	53.5	8.2	320	4
Foreign government nonprofit	55.7	355.3	22.2	13.8	5.3	16.3	71.3	540	2
Not sufficiently identified	17.5	9.5	0.0	0.0	1.0	6.5	1.5	36	0
Any support acknowledgment	1506.4	419.5	91.3	19.8	80.7	374.1	86.1	2578	10
References with no support identified	511.1	607.3	414.2	105.2	164.4	229.5	238.4	2270	31

*Combining NEI and Other NIH into a single category results in the following counts: US academic = 1198.7; foreign academic = 67.0; US for profit = 34.3; foreign for profit = 6.5; federal government = 58.2; other US = 289.2; other foreign = 16.2; total = 1670; no information = 4. These counts are less than the sum of the two rows in the table because of the avoidance of double counting when a single paper acknowledges both an NEI and a non-NEI NIH institute.

6. Examining the characteristics of agency-funded papers cited in patents

John Perko and Francis Narin, December 1996.

Objective

To examine the impact of agency-funded science on patents and to assess the extent to which domestic patents were citing agency supported work.

Methodology

From a data set of all papers with at least one US author and cited by US patents issued in 1993-94, papers supported by or authored at NIH were identified. Of the 223,000 US patents issued in 1993-94, 5,000 cited NIH-supported papers, of which 13,000 US-authored papers (cited within an 11-year window) were looked up in the library. There were almost 20,000 citations from patents to those papers, and 38,000 support acknowledgments (some NIH, some not). Indicators of patents and papers were produced.

Results

About 80 percent of the papers and citations supported by NIH were purely extramural. Biomedical Research and Clinical Medicine journals held more than 90 percent of these papers. Comparing NIH-supported Biomedical papers to all Biomedical papers cited by these patents, roughly one-third of the papers received some NIH support. Papers with at least some NIH support accounted for about 53 percent of all patent citations to Biomedical papers (Figure IV-17). About 63% of both citations and cited papers are from universities, colleges, and medical schools. The NIH-supported papers deal in strongly fundamental science, especially when compared with all Biomedical papers published. This suggests that fundamental, not applied, science is most conducive to biomedical innovation. It is fundamental science that critically relies on public funding. The NIH-supported cited papers are also published in highly influential journals. Further, NIH-supported Biomedical papers are cited by patents more than twice as often as all Biomedical papers published in the world. NIH-supported science is much more influential in patented technology than Biomedical science overall.

Finally, more than 70 percent of the patents are held by U.S. organizations employing U.S. inventors. The science that NIH supports stimulates technology that stays in domestic hands.

Comments

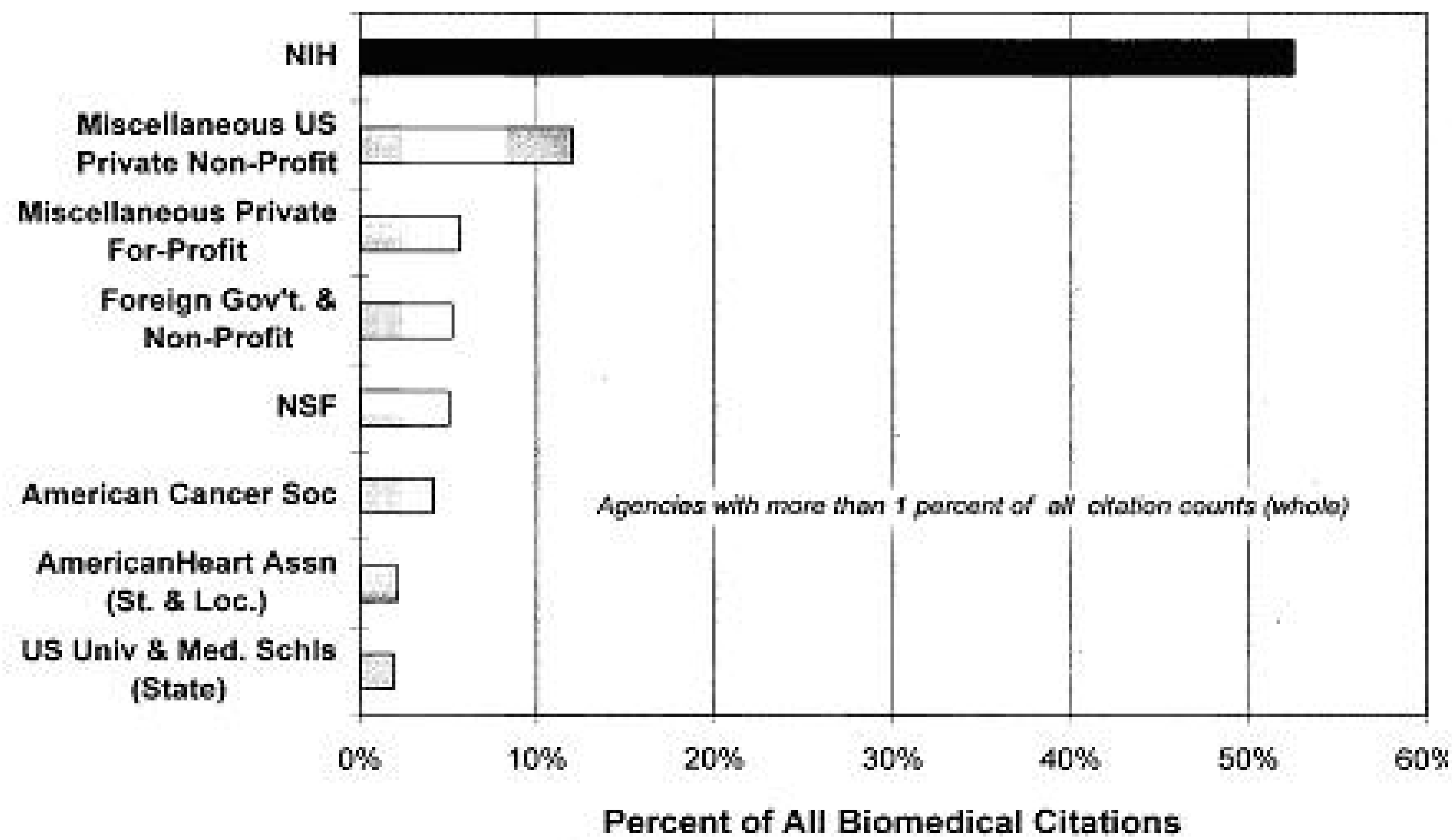
A related study of NIH-sponsored science and non-Biomedical patents was conducted in parallel with this one.

(CH435)

Figure IV-17.

Top Funding Agencies Acknowledged on All Biomedical Papers Cited by Patents

1993-94 Patents Citing U.S.-Authored Biomedical Papers Published in 1981-91



Includes only papers with explicit support acknowledgements

7. Examining the importance of NIH-funded research to recently approved drugs

Peter Kroll and Francis Narin, January, 1997.

Objective

This case study was one in a series that looked back from important technologies to identify the science underpinning them and beyond that to the agencies that funded the research. The study reviewed here looked at commercially important medical technology, as represented by drugs recently approved by the US Food and Drug Administration. These drugs were selected for study because they represent an important midway point in product development between the older, already approved pharmaceuticals and future pharmaceutical products for which patents have just been issued. This study's objective was to trace this set of drugs back to their basic scientific research underpinnings (and the funding support for the research) using patent information. The "top-down" approach was taken, using information available on the drugs to determine the patents protecting their technology, then tracing the papers cited by those patents back to their original funding source. This method avoids having to know in advance of the study the chronology of the fundamental research breakthroughs for each drug from beginning to end, or having to perform a literature search of each drug's development.

Methodology

The set of drugs was identified in collaboration with a subject expert. The set comprised drugs approved during 1995 and part of 1996, and was compiled from "New Drug Approvals in 1995" (Pharmaceutical Research and Manufacturers of America, January 1996) and the FDA's Web page of "Approved Drug Products with Therapeutic Equivalence Evaluations" through approximately September 1996.

The patents associated with the newly approved drugs were determined from three sources: the patents listed on each drug's application to the FDA for approval, The Merck Index, and the IMS World Patents International database. Because of the aggregate survey nature of the study, a detailed exploration of each drug's scientific and technological literature was infeasible, although additional patents and papers would likely be identifiable.

The selection process resulted in (exactly) 100 unique patents. Five of these patents were dropped, having been issued before 1975 (the date from which point on the US patent database was in machine-readable form). The remaining 95 patents represented 48 of the 55 drugs in the original set. Of these, 26 of the drugs were related to patents that cited science papers. Those patents cited 307 nonpatent references that were looked up in academic libraries for funding and bibliographic information. A relational database was developed for the patents and the 117 papers found (representing 155 science references).

Results

Three of the top four funding sources were found to be NIH institutes. Of the 26 drugs represented by patents citing science papers, Figure IV-18A shows that 12 of the drugs had patents citing papers that acknowledged NIH support. Of the 117 papers that were found in library lookup, 24% contained an acknowledgment to NIH funding. Adding in the 4.5% of papers that indicated no NIH funding but were authored by intramural NIH scientists, fully 29% of the found papers can be traced to NIH support.

Another way of looking at the statistics is from the point of view of the citing patents in

Figure IV-18B. Of the 44 patents containing journal references, 18 of them cited at least one NIH-funded or -authored paper.

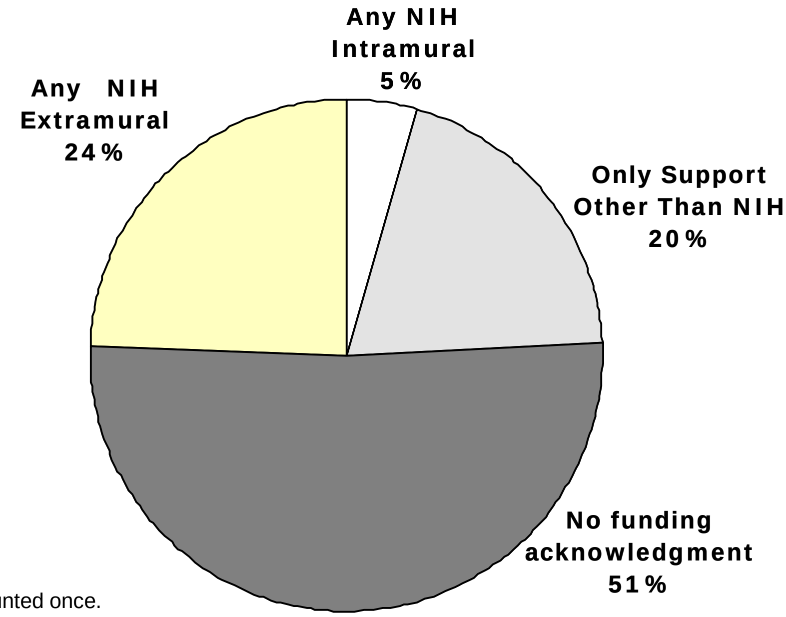
Comments

As found in other studies by CHI Research, the number of non-patent references cited increased dramatically over the years covered by these patents. Patents in this set from 1975-82 averaged 0.81 NPRs per patent; by 1990-06, the average was 4.89. Thus, the early patents provided few opportunities to trace related research funding.

(CH434)

Figure IV-18A.

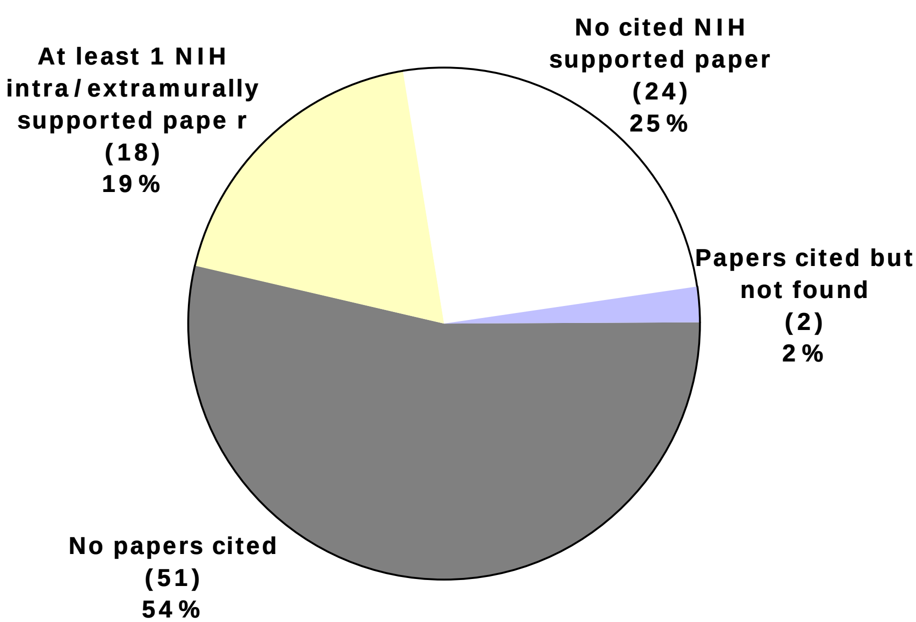
Support for Papers Cited in "FDA SET" Patents*



*Each paper counted once.

Figure IV-18B.

NIH Funding Support for Papers Cited in "FDA SET" Patent: Counted by Patent



8. The Linkage between Agency Sponsored Science and Patented Technology Outside its Core Area

John Perko and Francis Narin, January 1997

Objective

To measure the contributions that NIH (National Institute of Health) funded science makes to technologies outside biomedicine.

Methodology

A previous study had identified all U.S. patents issued in 1993-94 that referenced on their front pages journal articles that either explicitly acknowledged NIH support or listed at least one NIH address. This study eliminated from the patent set all medical patents. USPTO classifications were used to identify most medical patents and the titles and abstracts of the remaining patents were scanned to eliminate other medically related technologies. 242 non-medical patents citing 344 different papers formed the basis of this analysis.

Results

Plant science, pesticides and instruments and measurement together comprise about 60% of the non-medical patents. Food science and computing/numerical methods accounted for about 10-12% each. Other areas citing NIH research were composite materials & structures, cosmetics, hazardous waste and household textiles and cleaning. About 75% of the citations went to papers in biomedical research and clinical medicine, the core of NIH research. Other scientific areas were also cited - see Figure IV-19.

The papers cited by non-medical patents are as or slightly more fundamental science than all NIH-supported papers cited in 1993-94 patents. 80% of the citing patents were held by U.S. organizations establishing that the science NIH funds is largely used in domestic technology.

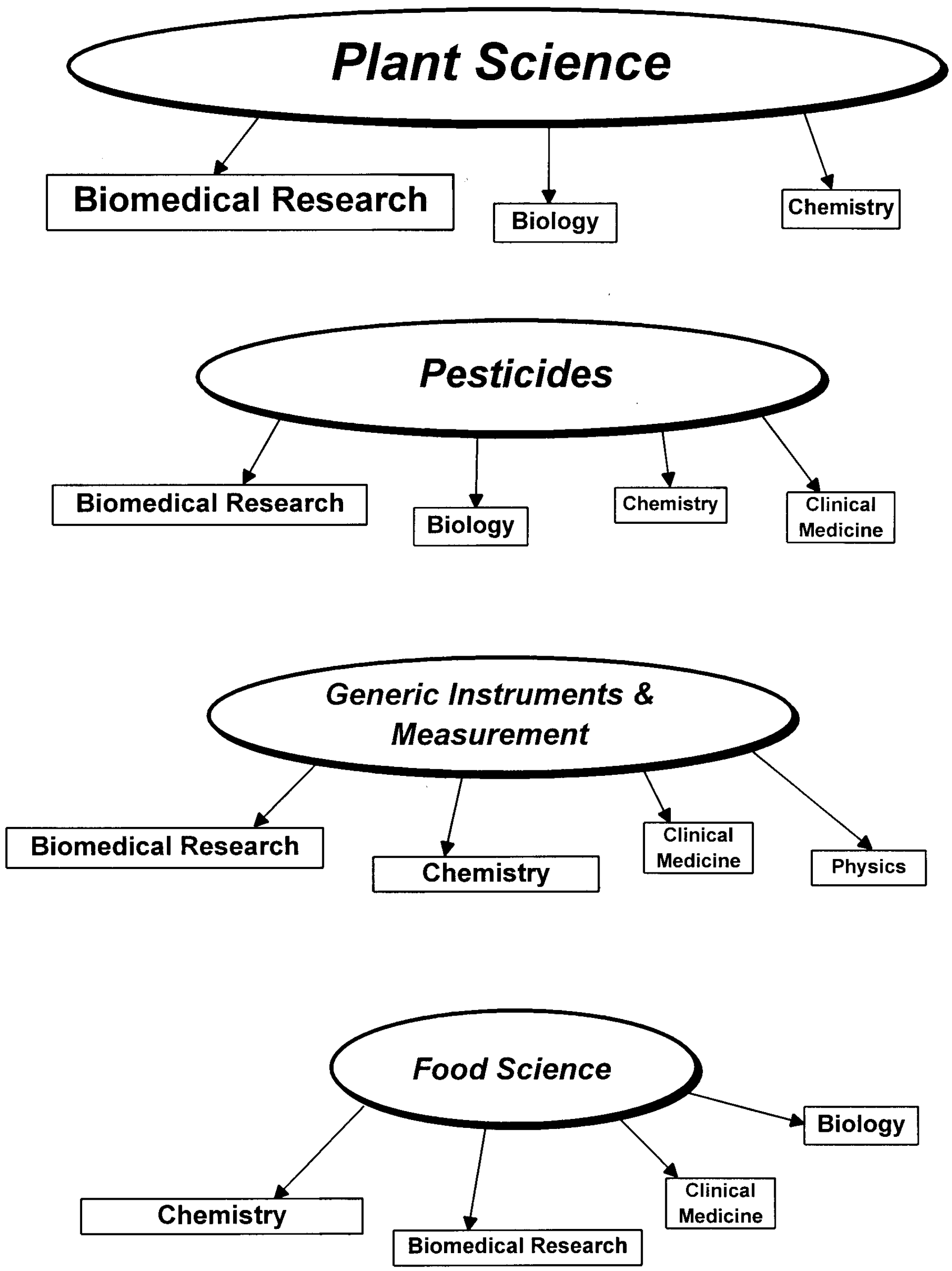
Comment

Theory tells us that scientific research has broad and unforeseen impacts. This study mapped the technologies in which NIH science had an impact outside its core medical area. In large part, it was NIH's core biomedical science that was drawn upon, not science peripheral to its mission. Demonstrating the importance of domestic users of the agency's science can be important if politicians are concerned that they are funding science and foreign technology benefits.

(CH437)

Figure IV-19.

Non-Medical Patent Technologies Depending on NIH Science



9. Examining the economic impact of technologies whose science base can be traced to NEI-funded research

Peter Kroll and Francis Narin, December, 1997

Objective

To assess that impact of National Eye Institute-funded research on two selected technologies. One is summarized here.

Methodology

The drug latanaprost is a prostaglandin-based treatment for reducing intraocular pressure. Thirty-five patents issued between 1975-94 were identified in an earlier study, and 77 papers cited by those patents were found in the library. Of these, 31 showed no funding support listed. Of the remaining 46 papers, 33 acknowledged National Eye Institute support, an additional 4 showed support by other NIH sources, and the remaining 9 were supported without NIH funds.

Results

Figure IV-20A shows the linkage between patents and papers in this area. Patents are represented at the top of the page, papers at the bottom. NEI-funded papers are indicated by a solid rectangle, others by an empty one. Lines are drawn to each NEI-funded paper from each patent citing it (represented by a solid oval). Cites to non-NEI-funded papers are not drawn, to prevent obscuring the cites to NEI-funded papers. The density of the citation traffic shows graphically the influence of NEI-funded research.

Fourteen of the 35 patents cited a total of 12 different papers by Laszlo Bito of Columbia University, a researcher supported steadily by NEI grants. Figure IV-20B shows the linkages to Bito's papers alone, illustrating this key researcher's seminal influence on the technology.

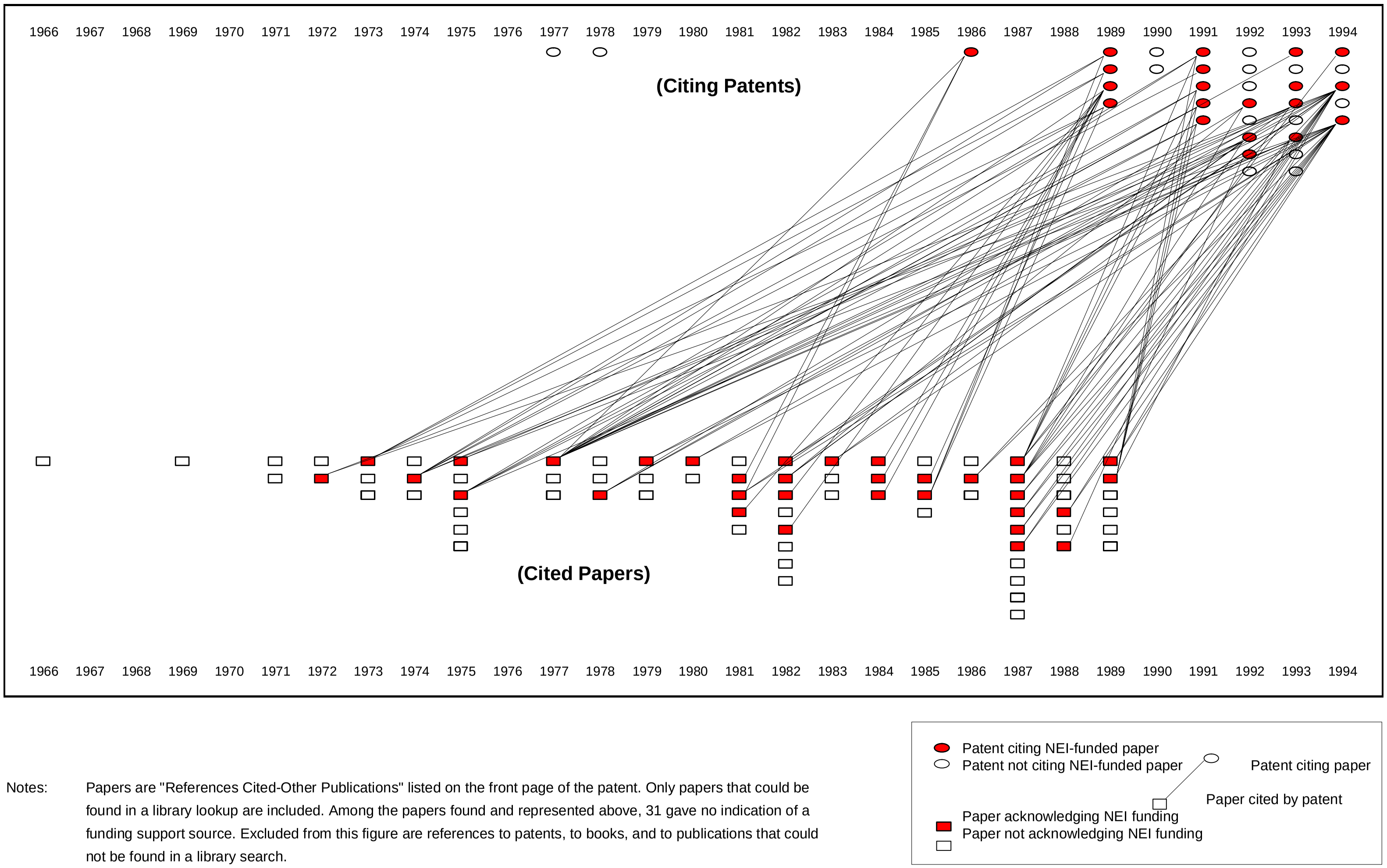
Comments

Pharmacia and Upjohn began marketing latanaprost as Xalatan in September 1996 with $9.8 million in sales. Subsequent quarterly reports showed sales totaling $129 for the four quarters ending September 1997. While the precise fraction of commercial impact directly attributable to NEI funding is impossible to determine from these data, the evidence presented in this brief study indicates a significant influence. To trace any individual funder's, scientist's, inventor's, or company's contribution as a fraction of economic output would require an extensive detailed inquiry into the history of the product from basic scientific exploration to final product engineering and marketing. But an approach taken in this study does demonstrate a significant impact by certain actors.

(CH460)

Figure IV-20A

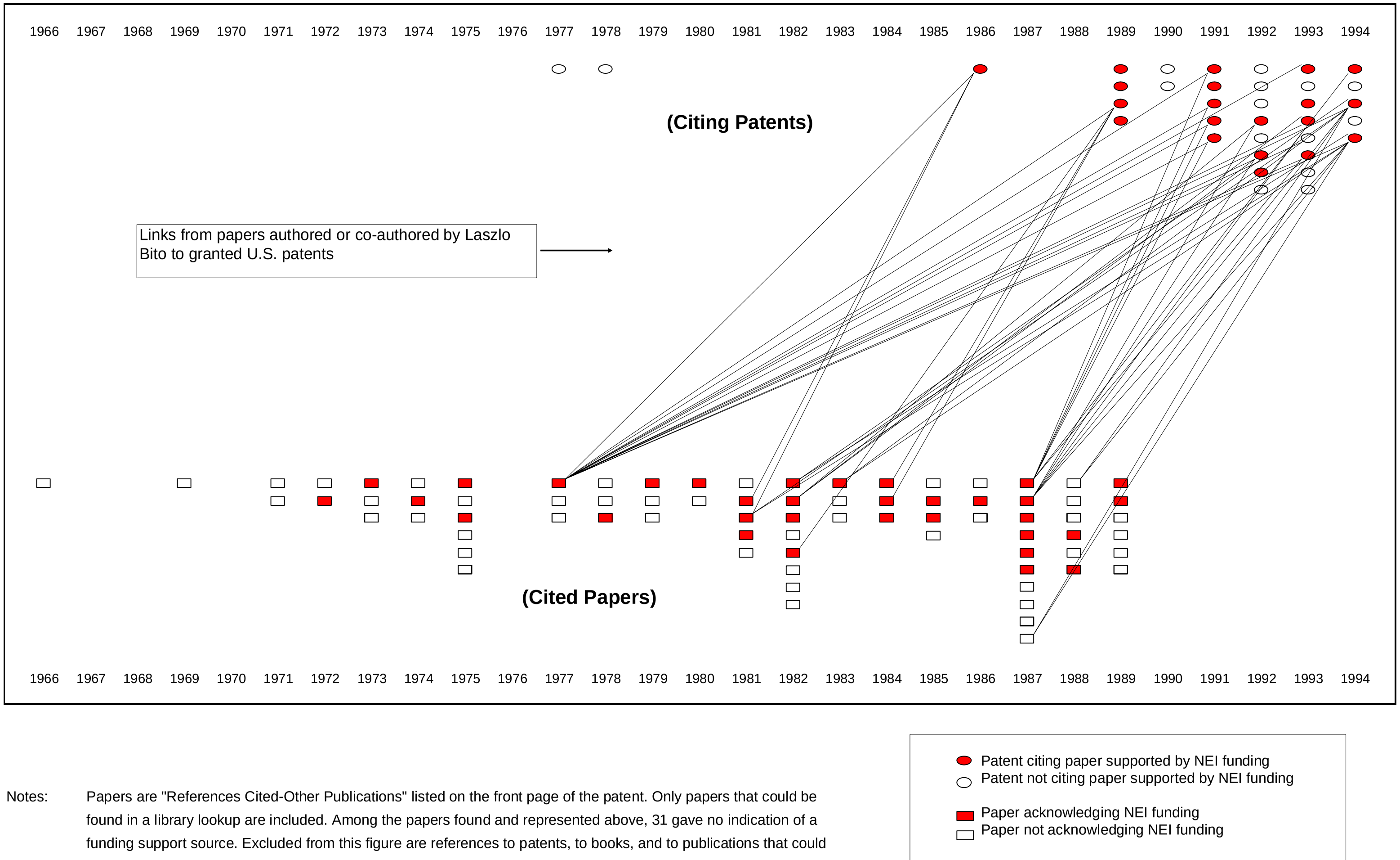
Linking Between Prostaglamdin Glaucoma Technology and NEI-Funded Research



Notes: Papers are "References Cited-Other Publications" listed on the front page of the patent. Only papers that could be found in a library lookup are included. Among the papers found and represented above, 31 gave no indication of a funding support source. Excluded from this figure are references to patents, to books, and to publications that could not be found in a library search.

Figure IV-20B.

Bito's Role in Science Literature and Technology of Prostaglandins in Treatment of Glaucoma



Notes: Papers are "References Cited-Other Publications" listed on the front page of the patent. Only papers that could be found in a library lookup are included. Among the papers found and represented above, 31 gave no indication of a funding support source. Excluded from this figure are references to patents, to books, and to publications that could not be found in a library search.

10. Examining the economic impact of spin-off technologies whose science base can be traced to NEI-funded research

Peter Kroll and Francis Narin, December, 1997

Objective

To evaluate the impact of National Eye Institute-funded research on technologies not directly related to eyecare.

Methodology

By using the previously-compiled set of NEI-funded papers cited by eyecare technology patents (derived from earlier studies), all 525 patents issued from 1983-96 citing those papers were then identified, including 390 eyecare patents already identified. In addition, all US-authored papers cited by patents issued from 1987-88 and 1993-94 had been looked up for funding sources in an earlier non-NEI study so other NEI-funded papers not cited by eyecare technology patents were determined for those patent years. After accounting for overlaps in the sets and removing a handful of eyecare patents newly identified in these sets, a final set of 262 non-eyecare related patents were identified for further examination.

Results

Of the 679 patents citing NEI papers in these sets, 39% were non-eyecare-related. Of these "spin-off" patents, the top three assignees citing to NEI were universities; the next ten were corporations.

The spin-off patents were primarily in other biomedical fields. They were further categorized by technology and application area, shown in Figure IV-21. Drugs and basic science/research represented the bulk of technology categories. Diagnosis and medical/surgical technologies were also well-represented. Just six patents fell into the "other" category (such as a doze detector and digital advanced television systems).

Breaking down the patents by specific biomedical application area, more than half the patents reflected the most basic of biomedical science areas such as laboratory techniques that cannot be categorized by medical application. Patents applicable to cancer represented about one-fifth of the total, with wound/cutaneous, cardiac/circulatory, and infectious disease categories ranging from 10% to 12% of the patents.

Several of these spin-off patents are highly cited by other patents, potentially indicating important technologies.

Comments

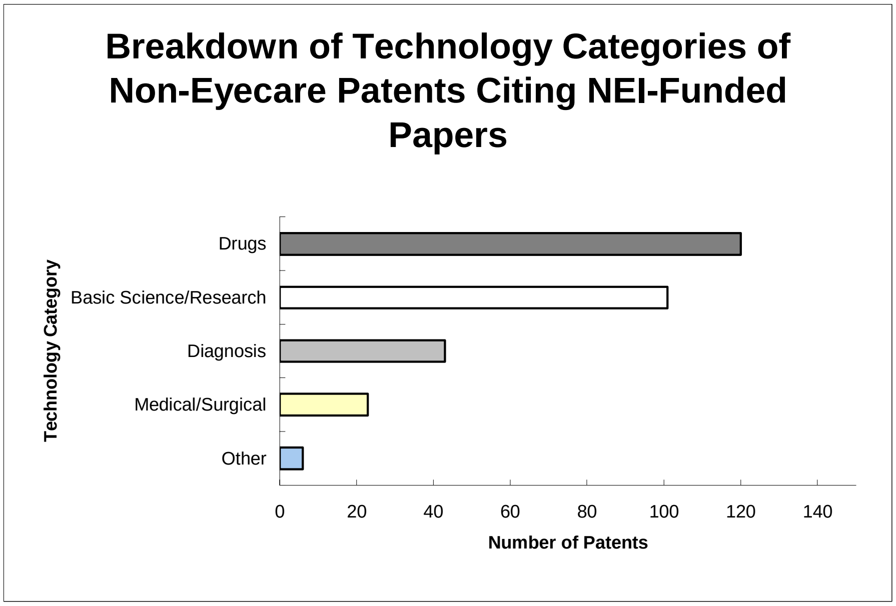
Though beyond the scope of this study, an analysis of the economic impact of these spin-off patents would give the insight of a quantifiable measure of impact of NEI-funded research beyond eyecare technology. These, coupled with the significant influence on non-eyecare medical-related patents show the broad reach of a single agency's research funding.

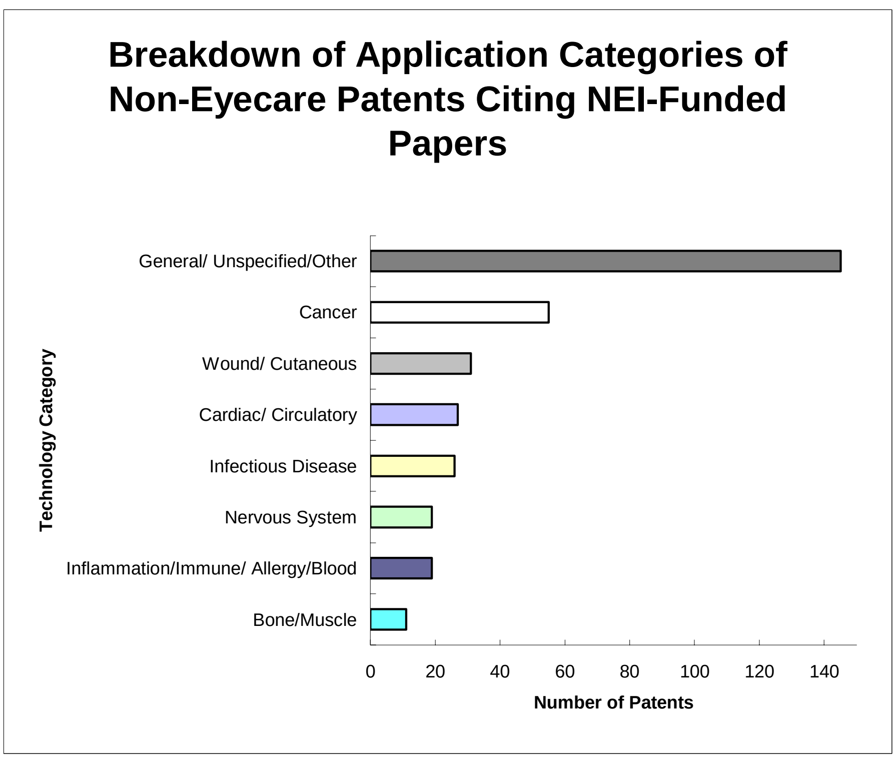
The funding identification characteristics were constrained by the available data sets. Specifically, this approach missed those patents in years other than 1987-88 and 1993-94 that cited a paper funded by NEI that had not been identified earlier as having been cited by an eyecare patent.

Another case of a patent that would not be found would be one citing a paper funded by NEI but with no US author-affiliated institution.

Figure IV-21.

Distribution by Categories





11. Examining the importance of publicly-funded research to STTR technology

Peter Kroll, Grace Ault, and Francis Narin, "Tracing the Influence of Basic Scientific Research on biotechnology Patents: A Case Study of Signal Transduction and Transcriptional Regulation (STTR)," Patent World, issue 100, March 1998, 38-46.

Objective

This case study was one in a series that looked back from important technologies to identify the science underpinning them and beyond that to the agencies that funded the research. The study reviewed here looked at research in the area of Signal Transduction and Transcriptional Regulation (STTR), one of the most promising arenas for new medicine development. Major diseases such as cancer, cardiovascular disease, and inflammatory conditions such as arthritis and asthma all involve abnormalities in signaling pathways that control cell growth or activity. Of interest was the origin--specifically funding--of research cited by STTR patents.

Methodology

200 1991-96 STTR patents were identified by keyword and company name search. Because of the extraordinarily high number of non-patent references cited by each, the sample was restricted to 1995 patents. The 62 patents each had an average of 26.2 non-patent references (a total of 1627 references). Upon further examination, 1,457 of these were references to journal science, and represented 1,361 different papers (some references cited the same paper).

Results

Patent ownership was widely dispersed among assignees, with 54 different organizations (some as co-assignees) owning the 62 patents. More than half of the assignees were private, for-profit corporations, suggesting this was an area of strong interest in commercial development. Author institutions included some pharmaceutical companies among the universities and research institutions. NIH institutes were mentioned as the authors' institutions 92 times, with Harvard Medical School accounting for 90. Publications in the field were found in journals with high influence scores and a very basic orientation, with prestigious journals dominating the list. Funding was dominated by NIH: the top three, five of the top six, and eight of the top eleven funding organizations were NIH institutes. Figure IV-22A shows that of all 1,297 papers that were found in the library, 49% contained an acknowledgment to NIH funding. Adding in the 5% of papers that indicated no NIH funding but were authored by intramural NIH scientists, fully 54% of the papers cited in the STTR patents could be traced to NIH support.

Another way of looking at the statistics is from the point of view of the citing patents, - Figure IV-22B. Of the 60 patents containing journal references, 54 of them cited at least one NIH-funded or -authored paper. Finally, in comparing indicator measures between the set of NIH and non-NIH papers, NIH-supported papers were published in more basic and influential subfields, and even comparing within subfields, were more basic and more influential than non-NIH papers

Comments

The area of STTR was not mature enough to produce an economic impact estimate of commercial products, which is consistent with the balance of public-private patent ownership in the area. The centrality of NIH work was made clear.

(CH432)

Figure IV-22A.

NIH Funding Support for Papers Cited in 1995 STTR Patents: Counted by Paper*

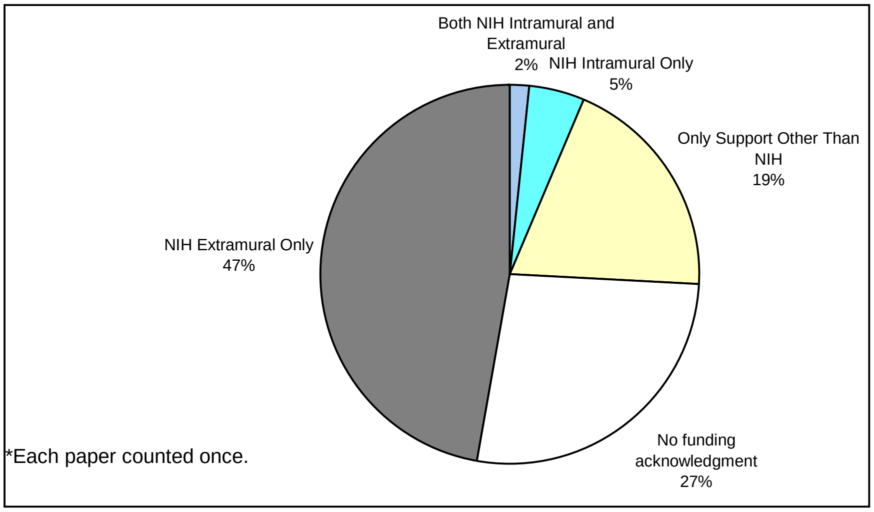
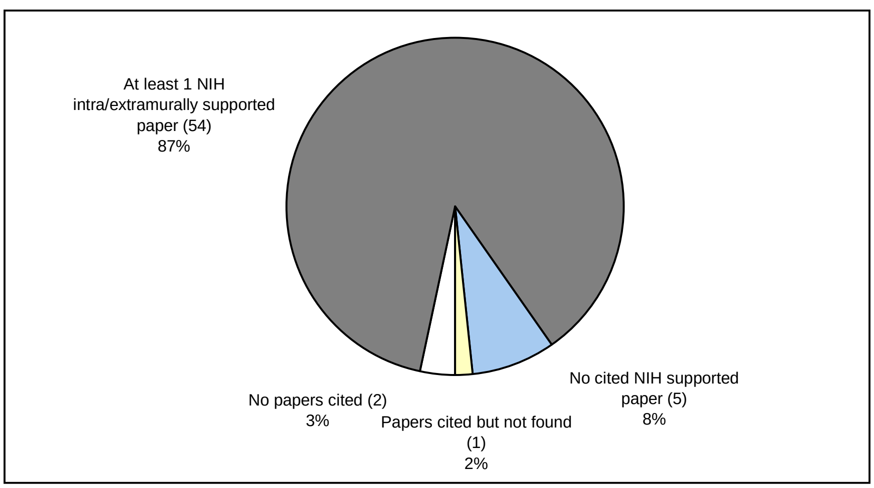


Figure IV-22A.

NIH Funding Support for Papers Cited in 1995 STTR Patents: Counted by Patent



12. Examining the characteristics of Australian government funding of papers cited in patents

Francis Narin, Michael Albert, Peter Kroll and Diana Hicks, February, 2000

Objective

The study examined the patterns in Australian patenting and the linkages between Australian patents and publicly supported science. The review presented here focuses on the relevant issues of Australian government (Australian Research Council and CSIRO) supported research.

Methodology

The study consisted of three aspects: linkages between Australian patents and the world's scientific research; linkages between world patents and Australia's scientific research; and the strengths and characteristics of Australian patents. In all cases, patents are only those issued by the US Patent and Trademark Office. The CHI Research patents-to-papers linkage database was used to analyze these areas. Two US organizations, NASA and Battelle Labs, were used to benchmark CSIRO patent indicators.

Results

The Australian government dominates patenting in Australia. CSIRO ranks as the leading Australian patenting enterprise, and non-CSIRO Australian Government patenting ranks second. There is little cross-sector citation among Australian patents (Figure IV-23A). Few Australian government patents are highly cited.

As for Australian science cited in patents, approximately 95 percent of all the citations to Australian science are to Australian public science: 63 percent to universities and their associated medical centers, 14 percent to CSIRO, 5 percent to private companies, and the rest to various other public sector institutions. Across all fields of research, papers from CSIRO and Australia's universities are prominent in the Australian science cited in patents. The Commonwealth of Australia (including CSIRO) ranks second in Australian patents (having at least one Australian inventor) that cite to Australian papers. Tabulations both of funding organizations and--for papers with no explicit funding acknowledgment--author institutions, show Australian government agencies in particular, and public institutions in general, as providing the greatest support of Australian science cited in patents. Compare the heavy citation of patents to public sector *papers* in Figure IV-23A to the low patent-to-patent citation in Figure IV-23B.

Comments

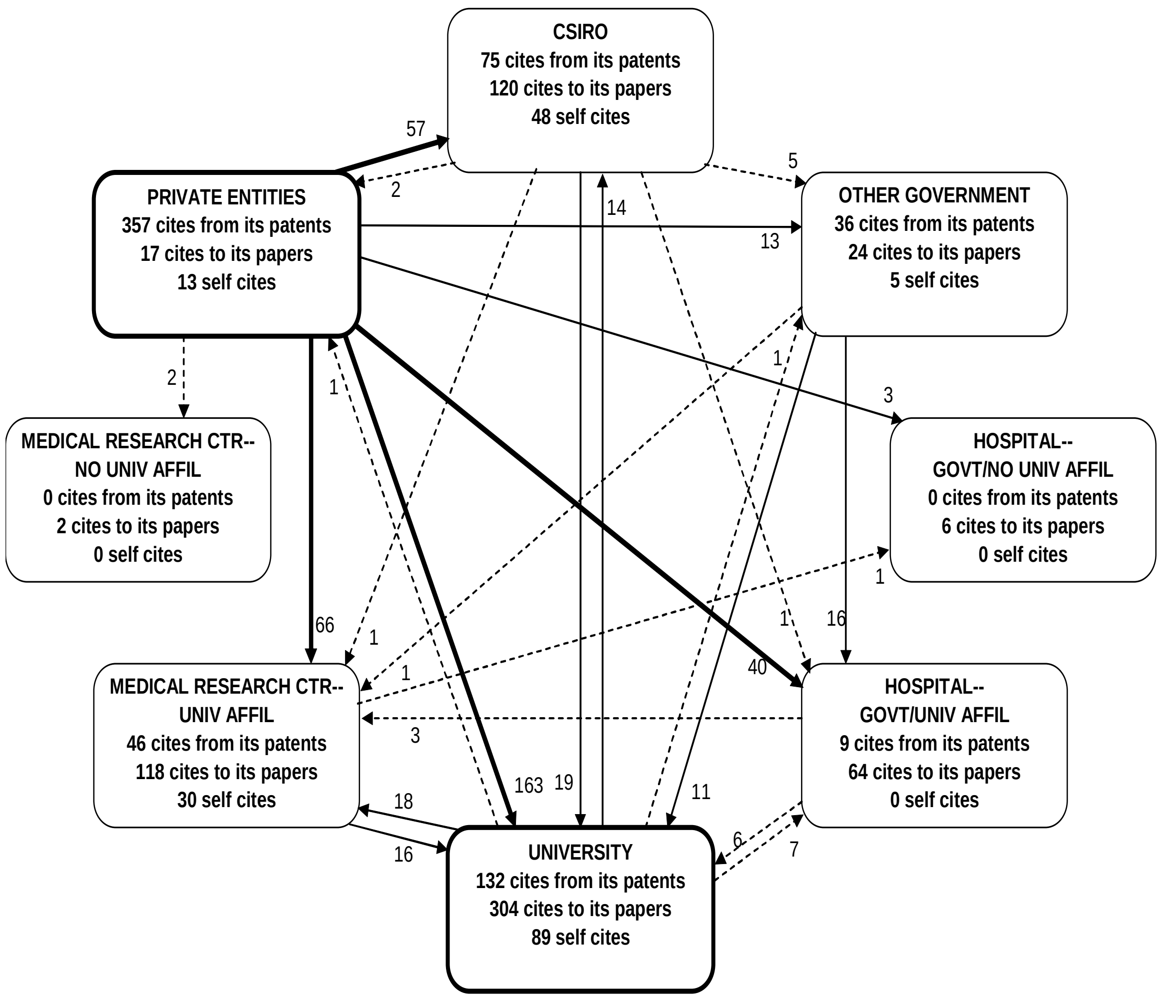
This study had a significant impact on public policy debate in Australia as it was released to coincide with national discussion on innovation. It is available on the Web at:

http://www.arc.gov.au/publications/arc_pubs/00_02.pdf.

NA

Figure IV-23A.

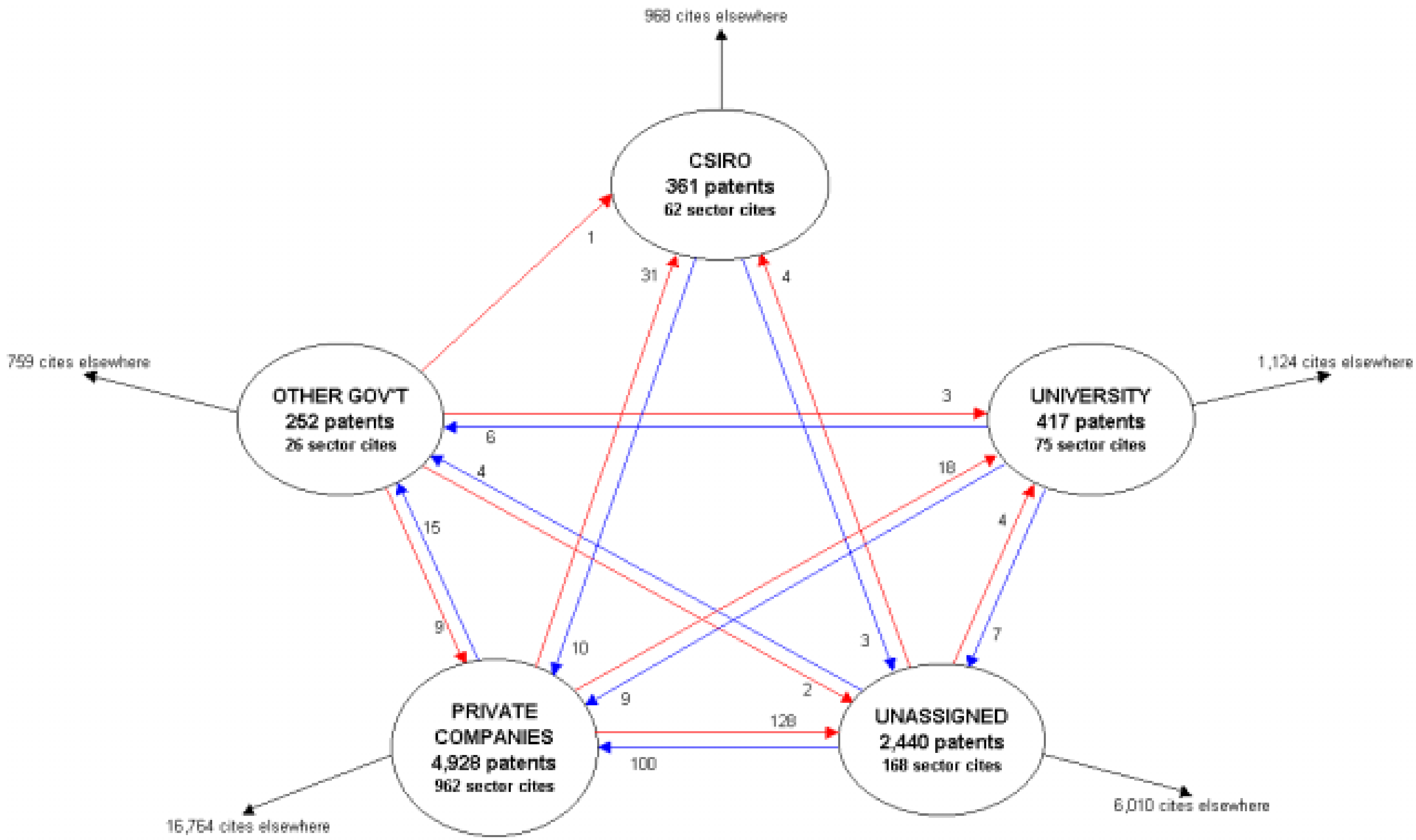
Citation of Papers by Patents Between Australian Sectors



Notes: Counts are of references by patents to papers.
Patents are counted fractionally; counts less than 0.5 have been omitted.
Patents by private entity inventors include foreign assignees.
Counts may not add up due to rounding

Figure IV-23B.

Cross-sector Citations Among Invented 1979-1998 U.S. Patents



Chapter V. References

(Items in bold are reproduced in the appendix)

Advisory Board for the Research Councils, The Royal Society, Economic and Social Research Council. *Evaluation of National Performance in Basic Research*, Department of Education and Science, London, 1986.

Anderson, Richard C., Francis Narin, and Paul R. McAllister. "Publication Ratings vs. Peer Ratings of Universities," *Journal of the American Society for Information Science, 29*, 91-103, 1978. Reprinted in Key Papers in Information Science, Belver C. Griffith, ed., Knowledge Industry Publications, Inc. 1980.

Averch, H.A. "Measuring the Cost-Efficiency of Basic Research Investment: Input-Output Approaches," *Journal of Policy Analysis and Management*, *6*, 3, 342-361, 1987.

Averch, H.A. "Exploring the Cost-Efficiency of Basic Research Funding in Chemistry," *Research Policy*, *18*, 165-172, 1989.

Averch, H. "Annotated Bibliography on Evaluation of Research, 1985-1990," in Bozeman & Melkers eds., 279-300, 1993.

Averch, H. "Criteria for Evaluating Research Projects and Portfolios," in Bozeman & Melkers, eds., 263-277, 1993.

Bozeman, B. & Julia Melkers, eds. *Evaluating R&D Impacts: Methods and Practice*, Kluwer, Dordrecht, 1993.

Buderi, Robert, John Carey, Neil Gross and Karen Lowry Miller. "Global Innovation: Who's in the Lead?" *Business Week Patent Scoreboard*, August 3, 1992.

Carpenter, Mark and Francis Narin. *Utilization of Scientific Literature by U.S. Patents*, final report to National Science Foundation on contract no. PRM-7801694, 1978.

Carpenter, Mark, M Cooper and Francis Narin. "Linkage Between Basic Research Literature and Patents." *Research Management*, *23*, 30, 1980.

Carpenter, Mark P., Francis Narin and Patricia Woolf. "Citation Rates to Technologically Important Patents," *World Patent Information, 4,* 160-163, 1981.

Carpenter, Mark & Francis Narin. "Validation study: Patent Citations as Indicators of Science and Foreign Dependence," *World Patent Information*, *5*, 180, 1983.

Carpenter, Mark. "Patent Citation as Indicators of Scientific and Technological Linkages." AAAS Annual Meeting. Detroit, MI, USA, May 30, 1983.

Collins, Peter & Suzanne Wyatt. "Citations in Patents to the Basic Research Literature" *Research Policy*, *17*, 65, 1988.

Committee on Science, Engineering, and Public Policy (COSEPUP). *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, National Academy Presss: Washington, D.C., 1999.

Coy, Peter, and John Carey. "The Global Patent Race Picks Up Speed," *Business Week Patent Scoreboard*, August 16, 1993.

Cozzens, S.E., Popper, S., Bonomo, J., Koizumi, K., Flanagan, A. *Methods for Evaluating Fundamental Science.* RAND/CTI DRU-875/2-CTI, Washington, DC, 1994.

Cozzens, S.E. *Assessment of Fundamental Science Programs in the Context of the Government Performance and Results Act GPRA*, Rand, Washington D.C., MR-707.0-OSTP, October 1995a.

Cozzens, S.E. "U.S. Research Assessment: Recent Developments", *Scientometrics*, *34 3*, 351-362, 1995b.

Deng, Zhen, Baruch Lev, and Francis Narin. "Science & Technology Indicators as Predictors of Stock Performance," *Financial Analysts Journal, 55,* 3, 20-32, May/June 1999.

Department of Defense. "Project Hindsight", Office of the Director of Defense Research and Engineering, DTIC Report No. AD495905, October 1969. See also, Sherwin, C.W. and R.S. Isenson, "Project hindsight: defense department study of the utility of research," *Science*, 156, 1571-1577, 1967.

Dietz, James S., Ivan Chompalov, Barry Bozeman, Eliesh O'Neil Lane, and Jongwon Park. "Using the Curriculum Vita to Study the Career Paths of Scientists and Engineers: An Exploratory Assessment," *Scientometrics, 49,* 3, 419-442, 2000.

Ellis, P., G. Hepburn and C. Openheim. "Studies on Patent Citation Networks," *Journal of Documentation, 34,* 1, 12-20, March 1978.

Garfield, Eugene. "Citation Indexes for Science," *Science, 122,* 108-111, 1955.

Geisler, E. *The Metrics of Science and Technology*, Quorum Books, Westport CN, 2000.

Georghiou, L. and D. Roessner. "Evaluating Technology Programs: Tools and Methods," *Research Policy*, *29*, 657-678, 2000.

Glanzel, W., A. Schubert and H.-J. Czerwon. "An Item-by-Item Subject Classification of Papers Published in Multidisciplinary and General Journals Using Reference Analysis," *Scientometrics*, *44*, 3, 427-440, 1999.

Griliches, Zvi. "Market Value, R&D and Patents". *Economics Letters, 7,* 183-187, 1981.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg. *Market Value and Patent Citations: A First Look*, Paper prepared for the Conference on Intangibles and Capital Markets, New York University, 1998.

Harhoff, Dietmar, Francis Narin, F. M. Scherer, and Katrin Vopel. "Citation Frequency and the Value of Patented Inventions," T*he Review of Economics & Statistics, 81*, 3, 511-515, August 1999.

Hicks, D. "The Difficulty of Achieving Full Coverage of International Social Science Literature and the Bibliometric Consequences," *Scientometrics*, *44*, 2, 193-215, 1999.

Hicks, D., A. Breitzman, K. Hamilton, F. Narin. "Research excellence and patented innovation," *Science and Public Policy*, *275*, 310-320, 2000.

Hodgdon, J.D., S.J. Fitzsimmons, L.C. Kerpelman, M.C. Jerrett, D. Deal, J. Weinberg. *Methods for the Strategic Evaluation of Research Programs: The State -of-the-Art - Annotated Bibliography*, report to National Science Foundation on contract No. PRA 8400688 Abt Associates Inc.: Cambridge MA, 1985.

Inhaber H. and K. Przednowek. "Quality of Research and the Nobel Prizes," *Social Studies of Science, 6,* 33-50, 1976.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson. "Geographic Localization of Knowledge Spillovers and Evidenced by Patent Citations," *Quarterly Journal of Economics, 108,* 3, August 1993.

Katz, J.S. and Diana Hicks. "How much is a collaboration worth? A calibrated bibliometric model," *Scientometrics*, *40*, 541-554, 1997.

Kerpelman, L.C. & S.J. Fitzsimmons. *Methods for the Strategic Evaluation of Research Programs: The State -of-the-Art - Analytical Overview Report*, report to National Science Foundation on contract No. PRA 8400688 Abt Associates Inc.: Cambridge MA, 1985.

Kleinberg, J.M. "Hubs, Authorities and Communities", *ACM Computing Surveys*, *31*, 4, December, 1999, available at: http://dev.acm.org/pubs/articles/journals/surveys/1999-31-4es/a5-kleinberg/a5-kleinberg.pdf.

Kostoff, R. "Evaluation of Proposed and Existing Accelerated Research Programs by the Office of Naval Research," *IEEE Transactions on Engineering Management*, *35*, 4, 1988.

Kostoff, R. "Evaluating Federal R&D in the United States", in Bozeman & Melkers, eds. 163-178, 1993.

Kostoff, R.N. "Federal Research Impact Assessment: State-of-the-Art," *Journal of the American Society for Information Science*, 456, 428-440, 1994.

Link, A.N. *Evaluating Public Sector Research and Development*, Praeger, Westport, CN, 1996.

Logsdon, J.M. & C.B. Rubin. "Research Evaluation Activities of Ten Federal Agencies," *Evaluation and Program Planning*, *11*, 1-11, 1988.

Martin, B.R. and J. Irvine. "Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy," *Research Policy*, *12*, 61-90, 1983.

McAllister, Paul R., Francis Narin and James G. Corrigan. "Programmatic Evaluation and Comparison Based on Standardized Citation Scores," *IEEE Transactions on Engineering Management*, *30*, 4, 205-211, November, 1983.

Melkers, J. *Bibliometrics as a Tool for Analysis of R&D Impacts*, in Bozeman & Melkers, eds. 42-61, 1993.

Melkers, J. and D. Roessner. "Politics and the Political Setting as an Influence on Evaluation Activities: National Research and Technology Policy Programs in the United States and Canada," *Evaluation and Program Planning*, *20* 1, 57-75, 1997.

Meyer, Martin. "Does Science Push Technology? Patents Citing Scientific Literature" *Research Policy*, *29*, 409-434, 2000.

Narin, Francis. *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*, Contract NSF C-627, National Science Foundation. March 31. Monograph: 456 NTIS Accession #PB252339/AS, 1976.

Narin, Francis and Elliot Noma. "Is Technology Becoming Science," *Scientometrics, 7,* 3, 369-381, 1985.

Narin, Francis, Elliot Noma and Ross Perry. "Patents as Indicators of Corporate Technological Strength," *Research Policy, 16,* 143-155, 1987.

Narin, Francis. Globalization of Research, Scholarly Information and Patents – Ten Year Trends. In: Proceedings of the North American Serials Interest Group NASIG 6th Annual Conference June 14-17, *The Serials Librarian, 21,* 2-3, 1991.

Narin, Francis & Dominic Olivastro. "Status report: Linkage Between Technology and Science," *Research Policy*, *21*, 237, 1992.

Narin, F. and K.S. Hamilton. "Bibliometric Performance Measures," *Scientometrics*, *36*, 3, 293-310, 1996.

Narin, Francis, Kimberly S Hamilton and Dominic Olivastro. "The Increasing Linkage between U.S. Technology and Public Science," *Research Policy*, *263*, 317-330, 1997. Reprinted in the *AAAS Science and Technology Yearbook*, 1998. **"Study Finds Public Science is Pillar of Industry", *New York Times* Science Times Section, Tuesday, May 13, 1997**.

Narin, Francis and Dominic Olivastro. "Linkage Between Patents and Papers: an Interim EPO/US Comparison," Prepared for *Proceedings of the Sixth Conference of the International Society for Scientometrics and Infometrics*, June 16-19, 1997, Jerusalem, Israel. Published in a special edition of *Scientometrics*, *41*, 1, 51-59, 1998.

National Science Foundation. *Science Indicators 1972,* Report of the National Science Board, 1973, and subsequent biennial *Science and Technology Indicators* reports.

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

Office of Technology Assessment. *Federally Funded Research: Decisions for a Decade*, OTA-SET-490 Washington, DC: U.S. Government Printing Office, May 1991.

Patent and Trademark Office, U.S. Department of Commerce. *Technology Assessment and Forecast,* Sixth Report, June 1976.

Pinski, Gabriel, and Francis Narin. "Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics," *Information Processing and Management*, 12 5 97-312, 1976.

Popper, S. W. *Economic Approaches to Measuring the Performance and Benefits of Fundamental Science*, Rand, Washington D.C., MR-708.0-OSTP, October 1995.

Reisner, P.   A Machine Stored Citation Index to Patent Literature Experimentation and Planning. In: *Proceedings of Automation and Scientific Communications Annual Meeting 1963*, H.P. Lunh ed.. American Documentation Institute, Washington, DC, 1965.

Research Value Mapping Project. *Qualitative-Quantitative Case Studies of Research Projects Funded by the Office of Basic Energy Sciences*, final report to the Office of basic Energy Sciences, Department of Energy, February 19, 1999.

Roessner, D. *Use of Quantitative Methods to Support Research Decisions in Business and Government*, in Bozeman & Melkers, eds. 179-205, 1993.

Roessner, D., B. Bozeman, I. Feller, C. Hill, N. Newman. *The Role of NSF's Support of Engineering in Enabling Technological Innovation*, first year final report for National Science Foundation SRI International: Arlington VA, January 1997.

Roessner, D., C. Ailes, I. Feller, L. Parker. "How Industry Benefits from NSF's Engineering Research Centers", *Research, Technology, Management*, *41*, 5, 40-44, 1998a.

Roessner, D., R. Carr, I. Feller, M. McGeary, N. Newman. *The Role of NSF's Support of Engineering in Enabling Technological Innovation: Phase II*, final report to National Science Foundation SRI International: Arlington VA, May 1998b.

Roessner, David. "Quantitative and Qualitative Methods and Measures in the Evaluation of Research," *Research Evaluation*, *8*, 2, August 2000.

Rogers, J.D. & B. Bozeman. " 'Knowledge Value Alliances': An Alternative to the R&D Project Focus in Evaluation," *Science, Technology and Human Values*, *26*, 1, 2001.

Patent and Trademark Office, U.S. Department of Commerce. *Manual of Patent Examining Procedures*, Section 904.02, 6th edition, 1995.

Seglen, P.O. "Why the Impact Factor of Journals Should not be Used for Evaluating Research," *British Medical Journal*, 314, 498-502, 1997.

Trajtenberg, Manual. "A Penny for your Quotes: Patent Citations and the Value of Innovations," *Rand Journal of Economics, 21,* 11, 1990.

Appendix: Annotated Bibliography

Abt Associates.

A list of their evaluations with short summaries can be found at: http://www.abtassociates.com/areas/science-projects.html. This consulting firm has conducted a large number of evaluations for federal agencies.

Averch, H.A. "Measuring the Cost-Efficiency of Basic Research Investment: Input-Output Approaches," *Journal of Policy Analysis and Management*, *6*, 3, 342-361, 1987.

Combines bibliometric and econometric approaches. In the author's words: Examines the strengths, weaknesses and assumptions of the Irvine-Martin approach to assessing the research productivity of institutions by counting their absolute numbers of citations and publications. Suggests that there would be payoff in looking at information return per dollar of agency investment. Carries out the first known example of an efficiency comparison for basic research in NSF's behavioural science program.

Averch, H.A. "Exploring the Cost-Efficiency of Basic Research Funding in Chemistry," *Research Policy*, *18*, 165-172, 1989.

In the author's words: the first econometric study of research funding efficiency by a federal agency. Uses citations per dollar of research investment as the measure of efficiency. Shows that efficiency is determined by highly interactive variables such as investigators' experience and the labor intensity of a project. Shows there is enormous variation in efficiency and that at least for the given sample, some factors thought to be important by the scientific community such as quality of department are not.

Averch, H. *Annotated Bibliography on Evaluation of Research, 1985-1990*, in Bozeman & Melkers, eds. 279-300, 1993.

Comments in bibliography focus particularly on issue of use of evaluations. Averch takes the position that non-use stems from nature of policy making process. Thus improved methods are seen as rhetorical tools to sell studies, but do not increase the use of results in policy making. Averch is critical of the fact that studies are used to demonstrate that policy makers did a good job, but are not used to change resource allocations. Raises the issue of cost effectiveness of evaluation, given its non-use.

Averch, H. *Criteria for Evaluating Research Projects and Portfolios*, in Bozeman & Melkers, eds. 263-277, 1993.

Combines an economist's very rational way of looking at things with a perceptive view of the realities of life in government. Interesting is the proposal of practical strategies, or "algorithms" for evaluating three types of research: basic research, policy research for use in public decision making and innovation research. Each strategy is stripped to its core elements (p. 268).

Bozeman, B. & Julia Melkers eds. *Evaluating R&D Impacts: Methods and Practice*, Kluwer, Dordrecht, 1993.

Out of print and so not easy to obtain. Authors take positions on why evaluation methods are not more used. Implicit but ever present in these discussions is their limitation to the U.S. Federal context. OTA 1986 and Abt studies from mid-1980's are heavily relied upon.

Committee on Science, Engineering, and Public Policy (COSEPUP). *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, National Academy Press: Washington, D.C., 1999.

A discussion of GPRA from the leading advocate and contractor in the peer review area.

Cozzens, S.E. *Assessment of Fundamental Science Programs in the Context of the Government Performance and Results Act (GPRA)*, (Rand, Washington D.C., October 1995a, MR-707.0-OSTP).

Can be ordered from: http://www.rand.org. There is no charge.

Explains the requirements of the GPRA and how they relate to program evaluation. GPRA enshrined the distinction between outputs and outcomes and stated that outcomes are of primary interest. GPRA is problematic for research because the framework is that agencies set goals and measure their progress toward them. This makes sense for research activity (number of grants awarded, number of research students supported etc.), but not for the use and application of research which is largely out of an agency's control.
The most balanced and sophisticated treatment of evaluation methods in the GPRA context.

Cozzens, S.E. "U.S. Research Assessment: Recent Developments", *Scientometrics*, *34*, 3, 351-362, 1995b.
This quick survey shows that quantitative evaluation studies are relatively rare in the United States. Most of the assessment of federal research programs is descriptive and expert judgement seems to be the favored method. GPRA requirements and the problems they have caused are discussed. Four agency pilot projects to develop acceptable indicators are described; the projects illustrate the underlying tension of agencies avoiding quantification and the Office of Management and Budget demanding measurable indicators.

Department of Defense. "Project Hindsight," Office of the Director of Defense Research and Engineering, DTIC Report No. AD495905, October 1969. See also, Sherwin, C.W. and R.S. Isenson, "Project Hindsight: Defense Department Study of the Utility of Research". *Science*, *156*, 1571-1577, 1967.

Geisler, E. *The Metrics of Science and Technology*, Quorum Books, Westport CN, 2000.
There are chapters on: economic & financial, bibliometric, co-word, patents, peer review, and process outcomes (a variant of quantified case study), as well as chapters on application in industry, government, academia etc. Draws on a vast literature, covers a lot of ground and so wanders into general issues of science and technology policy. Takes an even handed, strengths and weaknesses approach.

Georghiou, L. and D. Roessner. "Evaluating Technology Programs: Tools and Methods," *Research Policy*, *29*, 657-678, 2000.
The most recent review of evaluation methods reports newer developments, and acknowledges that the literature on R&D evaluation is now too vast to be comprehensively reviewed. Roessner is an influential U.S. practitioner and the review introduces work with which he has been associated.

Hodgdon, J.D., S.J. Fitzsimmons, L.C. Kerpelman, M.C. Jerrett, D. Deal, J. Weinberg. *Methods for the Strategic Evaluation of Research Programs: The State -of-the-Art - Annotated Bibliography*, report to National Science Foundation on contract No. PRA 8400688, Abt Associates: Cambridge MA, 1985.
Bibliography of articles on bibliometrics, economic or stochastic decision models, peer review, studies using several methods and others. Two to three paragraphs of commentary on each study.

Kerpelman, L.C. & S.J. Fitzsimmons. *Methods for the Strategic Evaluation of Research Programs: The State -of-the-Art - Analytical Overview Report*, report to National Science Foundation on contract No. PRA 8400688, Abt Associates: Cambridge MA, 1985.
A report focusing on how methods were used, how they were combined in use and how effective they were. Concludes:
1) peer review is the "touchstone". Bibliometrics has demonstrated utility but studies recommend it be used in conjunction with other techniques (the same is not said of peer review). Econometric techniques are frequently propounded, but save for cost benefit analysis, most techniques have not received widespread currency.
2) Evaluation techniques are combined in various ways.
3) Quantitative techniques appear to be used most often on applied research. Peer judgements are often successfully quantified.
4) Various outcomes, impacts and prestige measures are used.
5) Formal, strategic evaluation of research programs is not regularly conducted in government or industry.
6) Lots of methods are proposed, some are used, none are replicated.

7) The impact of evaluation on planning and resource allocation is limited because policy often overrides the conclusions of evaluations when decisions are made.

Kostoff, R. "Evaluation of Proposed and Existing Accelerated Research Programs by the Office of Naval Research," *IEEE Transactions on Engineering Management*, *35*, 4, 1988.
Describes a method of evaluating a portfolio of agency projects using a peer review process. The process is standardized and fairly quantitative. Kostoff developed this process for the Department of Energy then transferred it to the ONR when he moved. The process is well regarded and was used more than once at DOE and ONR.

Kostoff, R. *Evaluating Federal R&D in the United States*, in Bozeman & Melkers, eds. 163-178, 1993.
A quick tour of main categories of method (peer review of grants, peer review of program outcomes, traces studies, bibliometrics, cost-benefit), with examples of peer review being the fullest, and the bibliometrics based on Narin's work. Nice overview on prevalence of techniques versus used, published studies. Lack of use discussed - relies on Averch.

Kostoff, R.N. "Federal Research Impact Assessment: State-of-the-Art," *Journal of the American Society for Information Science*, *45*, 6, 428-440, 1994.
Material in this paper will also be found in Kostoff's *Handbook*.
Fairly standard. Author has some new critiques from his close-up contact with scientists in his work at the Navy. Contains an interesting calculation of the full costs of a peer review. Proposes that the future lies with large databases and co-occurrence techniques.

Kostoff, R.N. *The Handbook of Research Impact Assessment*, Seventh Edition, DTIC Report Number ADA296021, Office of Naval Research: Arlington, VA, 1997.
This and other work by Kostoff is at: http://www.dtic.mil/dtic/kostoff/index.html
The Handbook is the central work by Kostoff as most of his publications use material taken from this source. It is notable for its massive bibliography. It also contains extensive discussions of evaluations undertaken by Kostoff for the Navy, most of which seem to be peer review based. Under bibliometrics it contains extensive discussion of 1) Kostoff's work analyzing documents using word co-occurrences and 2) his work identifying direct and indirect impacts of research on other fields and application areas using a network analysis based on survey data.

Link, A.N. *Evaluating Public Sector Research and Development*, Praeger, Westport, CN, 1996.
Describes evaluations Link conducted for ATP. Chapter 2 covers evaluation issues: the evaluation method should be . . . Alternative economic methods - what they are and some of their problems - are discussed: multiple programming analysis, economic impact analysis and benefit-cost analysis (the later in the most depth). Exhibits a practioner's sophisticated and detailed awareness of methodological problems.

Logsdon, J.M. & C.B. Rubin. "Research Evaluation Activities of Ten Federal Agencies," *Evaluation and Program Planning*, *11*, 1-11, 1988.
Based on 44 interviews with research managers in the 10 major federal research funding agencies. Concludes that few federal agencies were making substantial use of research evaluation approaches other than informed peer judgement and review. An innovative and structured peer review is discussed as is the use of bibliometric approaches to research evaluation by NIH, and impact studies carried out by several agencies. The role of the National Academy of Sciences in promoting and undertaking peer review evaluations is clear.

McAllister, P.R., F. Narin & J.G. Corrigan. "Programmatic Evaluation and Comparison Based on Standardized Citation Scores", *IEEE Transactions on Engineering Management*, *30*, 4, 205-211, 1983.
Derives a technique to create normalized citation scores comparable across fields using standard statistical methods.

Melkers, J. *Bibliometrics as a Tool for Analysis of R&D Impacts*, in Bozeman & Melkers, 42-61, 1993.

An even handed survey of bibliometric literature from someone without strong commitments for or against the method.

Melkers, J. and D. Roessner. "Politics and the Political Setting as an Influence on Evaluation Activities: National Research and Technology Policy Programs in the United States and Canada," *Evaluation and Program Planning*, *20*, 1, 57-75, 1997.

A wide ranging review of research evaluation seen with the context of general program evaluation and with a comparative perspective that brings the U.S. situation into sharper focus. Explores how the fragmented U.S. political system creates a fragmented situation for program evaluation. Describes evaluation at the Office of Basic Energy Sciences (BES of DOE); NIST's Manufacturing Technology Centers, NSF's Engineering Research Centers, and NSF's Small Business Innovation Research program (SBIR).

Narin, F. and K.S. Hamilton. "Bibliometric Performance Measures," *Scientometrics*, *36*, 3, 293-310, 1996.

The application of literature, patent and linkage bibliometrics to government research evaluation questions with reference to GPRA. Complements rather than repeats this report and is included in the appendix.

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

Available at: http://www.wws.princeton.edu/~ota/disk2/1986/8622_n.html

A classic. Perhaps the best review of economic and bibliometric measures and their use in the United States. Discusses the results of many studies, reaches measured, balanced and accurate conclusions and places the discussion in the broader policy context.

Office of Technology Assessment. *Federally Funded Research: Decisions for a Decade*, OTA-SET-490, Washington, DC: U.S. Government Printing Office, May 1991.

Available at: http://www.wws.princeton.edu/~ota/

Contains a five page discussion of evaluation which seems to rely on Chubin and Averch as contractors. Most of the discussion concerns practice in other countries.

Popper, S. W. *Economic Approaches to Measuring the Performance and Benefits of Fundamental Science*, Rand, Washington D.C., MR-708.0-OSTP, October 1995.

Can be ordered from: http://www.rand.org. There is no charge.

Discusses methods to measure the economic and social returns from investment in fundamental research. Bibliometrics are addressed, but of course bibliometrics does not measure economic returns. Report is a very good introduction to economic approaches. It offers a sophisticated, in depth discussion of several studies with careful consideration of strengths and weaknesses.

Research Value Mapping Project. *Qualitative-Quantitative Case Studies of Research Projects Funded by the Office of Basic Energy Sciences*, final report to the Office of Basic Energy Sciences, Department of Energy, February 19, 1999.

This and related papers not listed here can be obtained from:
http://rvm.pp.gatech.edu/papers/index.html. Juan & Rogers, 2000 is related.

An evaluation based on 28 case studies. The evaluation was quantitative in that answers to standardized questions were quantified and statistical methods were used to find patterns of program characteristics associated with success. Four lessons were learned about the BES program. 1) Stable BES funding is invaluable to the field. 2) New organizational designs contribute to project effectiveness. 3) Basic research often spills over, resulting in new directions. 4) Interdisciplinary work requires a different management concept. 5) University-based projects differ little from those at federal labs.

Roessner, D. *Use of Quantitative Methods to Support Research Decisions in Business and Government*, in Bozeman & Melkers, 179-205, 1993.

One paragraph descriptions of studies undertaken for US federal agencies. Attributes lack of use to weakness in methods.

Roessner, D., B. Bozeman, I. Feller, C. Hill, N. Newman. *The Role of NSF's Support of Engineering in Enabling Technological Innovation*, first year final report for National Science Foundation, SRI International: Arlington VA, January 1997.

Available at: http://www.sri.com/policy/stp/techin/

Roessner, D., C. Ailes, I. Feller, L. Parker. "How Industry Benefits from NSF's Engineering Research Centers", *Research, Technology, Management*, *41*, 5, 40-44, 1998a.

Report of a survey of companies involved with university research centers. Assesses their expectations against the benefits they gained.

Roessner, D., R. Carr, I. Feller, M. McGeary, N. Newman. *The Role of NSF's Support of Engineering in Enabling Technological Innovation: Phase II*, final report to National Science Foundation, SRI International: Arlington VA, May 1998b.

Available at: http://www.sri.com/policy/stp/techin2/

Roessner, David. "Quantitative and qualitative methods and measures in the evaluation of research," *Research Evaluation*, *8*, 2, August 2000.

In the author's opinion, quantitative indicators pose dangers. Examples concerning evaluation of programs seeking to encourage collaboration are given.

Rogers, J.D. & B. Bozeman. "'Knowledge Value Alliances': An Alternative to the R&D Project Focus in Evaluation," *Science, Technology and Human Values*, *26*, 1, 2001.

There are more related papers than are listed here. See Research Value Mapping, 1999 for URL to obtain related papers.

A presentation of the rationale behind the Research Value Mapping and a development of theory. Rogers & Bozeman propose a new approach to research evaluation. The ideas arose out of an evaluation for the Department of Energy. He argues that projects and programs are most often bureacratic fictions that bear little relation to the nature of research work. He concludes that evaluation should move closer to research as it is conducted by taking as its unit of analysis the alliances of researchers and others that work on a research topic and its application. The article discusses these ideas and several examples of networks of different types. The approach is well grounded in the most advanced sociology of science and addresses the complex intermingling of fundamental and application activities and institutions in research today. This complexity has come to forfront in the decades since research evaluation methods were first developed. If further work can establish the characteristics of alliances that are associated with economic and social benefits, the ideas may serve to move evaluators into a more realistic treatment of the nature of research today.