Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research

Jeffrey L. Furman
Boston University

Scott Stern
Northwestern University, Brookings, and NBER

July 2002

Version 1.0

* This paper was prepared for the NBER Summer Institute, July 2002. We thank each of the scientists who graciously offered their insights, the personnel of the American Type Culture Collection, and especially Dr. Raymond Cypess and Robert Hay. Useful comments were provided by Tim Bresnahan, Shane Greenstein, Robert Litan, and Paul Romer. Jason Corradini, Mercedes Delgado, Lorraine DeLeon, Chijoke Emineke, Anna Harrington, Kinga Piekos, Martha Kam, Julia Lo, Megan MacGarvie provided excellent research assistance. All errors are our own. Financial support for this research was provided by the Brookings Institution. Author contact information: Scott Stern, Kellogg Graduate School of Management, 2001 Sheridan Road, Evanston, IL 60208 s-stern2@northwestern.edu, and Jeffrey Furman, Boston University School of Management, 595 Commonwealth Ave #639a, Boston, MA 02215, furman@bu.edu.

Climbing Atop the Shoulders of Giants: Identifying the Impact of Institutions on the Knowledge Cumulation

Jeffrey L. Furman
Scott Stern

ABSTRACT

While the cumulative nature of knowledge production has been recognized as central to the process of economic growth, the microeconomic and institutional foundations of cumulativeness are less understood. This paper disentangles two distinct mechanisms -- selection and marginal effects -- by which institutions impact cumulativeness. The selection effect results from the fact that institutions (e.g., prestigious universities) may be associated with high spillovers because the associated researchers (and the knowledge they discover) are of high intrinsic quality. The marginal impact of institutions is the increment to cumulativeness resulting from association with a specific institution (i.e., research may be more accessible because it is discovered in a university setting). This paper develops and implements an empirical framework distinguishing these effects in the context of a specific institution, biological resource centers (BRCs). BRCs are "living libraries" that authenticate, preserve, and offer independent access to biological materials, such as cells, cultures, and specimens. Relative to a "peer-to-peer" network of informal exchange, BRCs reduce the marginal cost to researchers of building on prior research efforts. The evaluation of how BRCs affects the cumulative impact of knowledge exploits three key features of the environment: (a) the impact of scientific knowledge is reflected in future scientific citations, (b) deposit into BRCs often occur with a lag after initial research is completed and published and (c) these "lagged" deposits are often the result of arguable (and testably) exogenous shocks. Employing a differences-in-differences estimator linking specific materials deposits to journal articles, we find evidence for both selection into and the marginal impact of BRCs. With article-specific fixed effects, the marginal impact of BRC deposit is estimated to increase the citation rate by 81%. Further, the marginal impact of biological resource centers increases with the "vintage" of the knowledge under consideration and has increased during the 1990s. Finally, a rate-of-return analysis suggests that, relative to traditional grant mechanisms, public expenditures towards authentication, preservation and access to research materials offers a three-fold gain in fostering the cumulativeness of scientific knowledge.

> "If I have been able to see further, it was only because I stood on the shoulder of giants."
>
> Isaac Newton, 1676

I. Introduction

Cumulative innovation is central to long-run economic growth. In order for growth to be sustainable, technologies developed and discoveries made at a point in time must serve as the building blocks for future research. If the knowledge pool stagnates, diminishing returns will set in and growth will halt. However, if the knowledge produced by each generation is built upon and serves to increase the stock of knowledge available to future generations, diminishing returns may be held at bay, allowing sustainable long-term growth (Romer, 1990). A distinctive feature of modern capitalism is that, across a wide range of industries and technologies, the process whereby researchers "stand on the shoulders of giants" seems to be self-perpetuating: from information technology to transportation to pharmaceuticals, technological and scientific productivity is maintained by researchers by drawing upon an ever-expanding set of knowledge applicable to their field.

Despite its apparent importance, the microeconomic origins and institutional foundations of cumulative innovation are not well understood. The conditions that allow innovation to be cumulative are subtle, since the mere production of a piece of knowledge does not at all guarantee that others will be able to exploit that piece of knowledge. At the very least, researchers must be aware of the existence of prior knowledge; in the absence of awareness, researchers must often "reinvent the wheel" to make further progress. More generally, the degree of cumulativeness depends on the costs researchers face to access and verify the fidelity of the prior knowledge used as the

basis for their research. Uncertainty about the robustness of prior findings necessitates a costly process of reverification and reinterpretation, reducing the productivity of current research efforts. Therefore, to be effective, cumulative innovation must somehow reduce these costs so that research productivity remains high even as researchers draw on an ever larger body of knowledge.

At a broad level, effective cumulative progress depends on the institutions, legal rules, and social structures which allow researchers to draw upon a "knowledge stock" when pursuing their own research. For example, the regional innovation system and the norms of the scientific system allow researchers to draw upon knowledge from their local region or scientific discipline (Nelson, 1993; Jaffe, et al, 1993; David and Dasgupta, 1994). At a more nuanced level, ability of a researcher to draw upon others' knowledge has been tied to their participation and position within the specific social network in which that knowledge is embedded (Powell, 1998; Rosenkopf and Tushman, 1998). Moreover, the ability to build upon the findings of previous researchers depends on the distribution of intellectual property rights and the potential for contracting between generations of researchers (Scotchmer, 1991).

The role of a number of specific institutions in facilitating knowledge spillovers is addressed by empirical research that often exploits bibliometric methods. This literature has demonstrated the broad impact and geographic diffusion attained by university research (Jaffe et al., 1993) and has elucidated aspects of the roles of patent policy (Mowery et al., 2001; Branstetter and Sakakibara, 2001), R&D consortia (Irwin and Klenouw, 1996), national laboratories (Jaffe and Lerner, 2001), venture capital (Kortum and Lerner, 2000), and patent pools (Lerner and Tirole, 2002) in contributing to knowledge spillovers.

In assessing the extent to which any institution influences the way in which the "knowledge stock" is created, maintained, and extended, researchers face a considerable challenge: Though conceptually distinct, it is empirically difficult to isolate the impact of a particular piece of knowledge from the institution in which it is embedded. We distinguish between two distinct mechanisms by which institutions impact the cumulativeness of knowledge. We refer to the first of these mechanisms as a selection effect. The selection effect acknowledges that institutions may be associated with high rates of knowledge spillovers because the researchers and knowledge associated with them are of higher intrinsic quality. Second, we define the marginal impact of an institution to be the increment to cumulative impact that knowledge of a given quality achieves by being associated with that institution. We develop in this paper a novel approach to disentangle the contribution of an institution from the qualities of the knowledge associated with it.

We focus on a specific type of institution, called biological resource centers (BRCs), which play an important role in life sciences research. BRCs collect, certify and distribute biological organisms for use in biological research and in the development of commercial products in the pharmaceutical, agricultural and biotechnology industries. BRCs maintain a large and varied collection of biological materials, including cell lines, micro-organisms, recombinant DNA material, biological media and reagents, and the information technology tools that allow researchers to access and exploit these biological materials. The ability to exploit prior research in the life sciences often depends on access to the cells, cultures, and specimens used in that research. Along with peer-to-peer distribution networks, for-profit companies that market biological materials, and private culture collections, biological resource centers constitute one of the institutional

arrangements by which scientists can obtain materials for research purposes. Our empirical analysis evaluates whether, relative to alternative institutional arrangements, the deposit of research materials in a particular biological resource center, the American Type Culture Collection (ATCC) is associated with knowledge having a greater impact on future research.

Our empirical approach builds on the recent studies that use citation analysis to investigate technological communities and the cumulativeness of discovery and innovation (Jaffe, et al, 1993; Griliches, 1998; Murray, 2001). Specifically, we exploit two facts in order to develop a differences-in-differences estimate of the impact of BRCs on knowledge spillovers. First, in most cases, each material deposited in a BRC is associated with a journal article describing its initial characterization and application. Second, various subsets of BRC deposits have been shifted exogenously from prior institutional arrangements into biological resource centers. For example, some collections that are maintained in a private university laboratory may be shifted into a public BRC if the principal investigator retires or switches universities. By comparing citation patterns between a sample of articles linked to BRC deposits with those of a control group (chosen as the preceding articles in the same issue of the same journal), we can ascertain whether knowledge associated with BRC materials has a greater than average impact on future research. This result may obtain, however, simply because researchers deposit materials that are intrinsically important. To distinguish this 'selection effect' from the marginal impact of the BRC on knowledge spillovers, we exploit the experiments associated with a few instances in which collections of materials were shifted exogenously into biological resource centers. By evaluating whether articles associated with such materials receive a 'boost' in citations (relative to the within-article

trend, controlling for the age of the article as well as time period effects), we obtain an estimate of the marginal impact of the BRC on knowledge cumulation.

Our principal findings demonstrate a dramatic impact of BRCs on citation patterns, both in the cross-section and in the differences-in-differences analysis. First, comparing BRC-deposited articles with a set of control articles, we find that articles associated with BRC deposit have a significantly higher rate of citation. Second, our differences-in-differences analysis provides specific estimates of the strengths of both the selection effect and marginal impact of BRCs. The estimates imply that articles deposited in BRCs receive approximately 180 percent more citations per year than control articles. Further, controlling fully for selection (via article fixed effects), as well as article "vintage" effects, and time effects, we observe a statistically significant and economically important effect of BRC-deposit on future citations. On average, BRC-deposit is associated with an approximately 80 percent boost in annual citations. For each of the exogenously shifted "special collections," except one transferred only a few years ago, the average post-deposit impact on annual citation ranges from between 50 percent to 100 percent. Further, the boost in citation experienced by BRC-linked articles is modest in the initial years after deposit, but grows substantially over time.

Using our estimates of the citation boost associated with BRC-deposits, our concluding analysis estimates a "rate-of-return" associated with depositing materials in biological resource centers. Benchmarking on estimates of the cost per academic citation of Adams and Griliches (1996) and on estimates of the accession costs of new BRC materials (OECD, 2001), we estimate that articles associated with deposits to biological resource centers achieve a nearly three-fold efficiency in terms of inducing citations relative to articles not associated with BRC deposits. Taken together, we interpret our

results as demonstrating that the impact of published scientific research on future scientific research depends on the institutional arrangement in which that knowledge is embedded.

The remainder of the paper proceeds as follows. Section II reviews antecedent research considering the role of institutions in cumulative knowledge spillovers. Section III describes biological resource centers, focusing on characteristics that may make them integral to cumulative research in the life sciences. Section IV outlines an empirical differences-in-differences framework for identifying the impact of the selection effect and the marginal impact of BRCs on knowledge spillovers. Sections V and VI review of the data and present the empirical results, respectively. A final section concludes.

II. The Role of Institutions in Cumulative Knowledge Spillovers

II.A. Cumulativeness and Institutions in Economic Growth and Knowledge Spillovers

The critical role of technological progress on economic growth has been appreciated at least since Solow (1957) and Abramovitz (1956). Economists in recent decades have focused even further on the link between sustained productivity growth and the vitality of sectors and industries with a strong connection to science and particular scientific disciplines (Rosenberg, 1974; Adams, 1990). Two critical features of the role of scientific advance and technological progress on economic progress are the cumulative nature of these processes and the importance of institutions in establishing the environment for invention and innovation and the mechanisms by these spillover across sectors and over time.

The cumulative nature of science and technical advance has been characterized famously by Isaac Newton in the phrase, "If I have been able to see further, it was only

because I stood on the shoulder of giants."[1] The role played by cumulativeness in economic growth is at the core of models of endogenous growth (Romer, 1990; Jones, 1995). In order to serve as a foundation for long-term growth, scientific research and technological progress must continually spill over across fields, economic sectors, and over time (Romer, 1990; Grossman and Helpman, 1991; Jones, 1995; Porter and Stern, 2000). In other words, in order to avoid diminishing returns to investments in ideas, research must continuously "stand on the shoulders" of prior knowledge.[2]

Economists have also long recognized that institutions are closely associated with the accumulation and diffusion of knowledge (Bush, 1945; Nelson, 1959; Rosenberg, 1963). Culled from the experiences of World War II, Vannevar Bush's *Science: The Endless Frontier* provided a clear and compelling articulation of the role that basic research funding and support for scientific progress could play in economy-wide prosperity and security (Bush, 1945). Nelson (1959) formalizes some of these ideas, describing the economic rationale that private investment tends towards technological innovation and arguing that public investments, in institutions such as universities, government laboratories, and other not-for-profit organizations, are needed to support scientific advance. Rosenberg (1963) emphasizes the role of institutions in affecting knowledge spillovers among related economic sectors and in determining the microeconomic environment for technical advance.

Building on these initial articulations of the importance of institutions to cumulative growth, economists over the past two decades have come to appreciate the

[1] This phrase was first used by Newton (1676) in a letter to Robert Hooke in the context of a dispute over the nature of light: "What Des-Cartes did was a good step. You have added much several ways, & especially in taking ye colours of thin plates unto philosophical consideration. If I have seen further it is by standing on ye sholders of Giants."

manner in which economic progress depends on the environment for innovation and the constellation of institutions supporting cumulative advance (Nelson, 1993). Particular consideration has been given to clarifying the roles of universities, scientific societies, patent offices, and archives in driving the diffusion of innovation and the process of knowledge spillovers among researchers over time (Mowery and Rosenberg, 1989; Rosenberg and Nelson, 1994; Dasgupta and David, 1994; Mansfield, 1995). Scotchmer (1991), for example, relates the structure in patent law to the microeconomic conditions fostering cumulativeness in technological innovation. Complementing this economic perspective on cumulative knowledge growth, a sociological perspective has emphasized that the ability of a researcher to draw upon others' knowledge is linked to their participation and position within the specific social network in which that knowledge is embedded (Powell, 1998; Rosenkopf and Tushman, 1998).

Recently, a sophisticated empirical literature has emerged that attempts to identify the impact of particular institutions on knowledge spillovers. This research often employs citations to academic papers or approved patents as a trace indicators of the influence of prior knowledge on current advances. A number of authors have focused on the extent of knowledge spillovers created by university research. For example, Jaffe et al. (1993) demonstrate that university patents receive citations at a significantly higher rate and with significantly greater geographical scope than an appropriate control group. Examining the impact of university science on commercial innovation, Branstetter (2000) reviews patterns of patent citations to academic research papers, finding that spillovers

[2] Further research in international economics examines the impact of knowledge spillovers across borders, accounting for and explaining its impact international trade patterns and economic growth across countries (Kortum and Eaton, 1996; Keller, 200; Coe and Helpman, 1995).

from academic science to commercialized inventions occurs in a limited set of technological fields and geographic areas.

Empirical research has also focused on the implications of patenting policies on knowledge growth and spillovers. For example, Mowery et al. (2001) and Mowery and Ziedonis (2002) find that the Bayh-Dole Act, which changed the U.S. policy with respect to the patenting of inventions funded by government grants, had a positive impact on the extent of patenting at three major research universities, but did not appreciably alter the content or "generality" of university patenting.[3] Additional research considers the impact of other institutions on knowledge growth and spillovers, including R&D consortia (Irwin and Klenouw, 1996; Branstetter and Sakakibara, 2001), national laboratories (Jaffe and Lerner, 2001), venture capital (Kortum and Lerner, 2000), and patent pools (Lerner and Tirole, 2002).

II..B. Institutions and Knowledge Spillovers: Selection versus Marginal Impacts

While extant research characterizes the impact of research and innovations in particular institutional on knowledge spillovers, it has not specifically disentangled the role played by institutions from the quality or match quality of the knowledge embedded within those institutions. For example, this research has not measured whether university patents are more highly cited because university technology tends to be more intrinsically important or because universities serve a crucial role in disseminating ideas upon which others can build.

[3] Branstetter and Sakakibara (2001) find that 1998 reforms that expanded the scope of Japanese patents has only a modest impact on R&D effort and innovative output. See, also, Kortum and Lerner (1999) regarding explanations for recent changes in patent output in the United States.

To disentangle this puzzle, a number of key barriers must be overcome. First, one must be able to define knowledge subject to spillovers whose diffusion can be tracked, whether the knowledge is associated with the institution under study or a control group. Second, these "pieces" of knowledge must be comparable to each other regardless of whether they are associated with the focus institution or are in the control group. Third, the assignment of knowledge to the institution or the control group must be exogenous. In other words, the drivers of the group in which pieces of knowledge are located should not be sensitive to the importance of that knowledge or the degree to which that knowledge might be useful for follow-on researchers.

In the remainder of this paper, we develop an empirical framework to isolate the marginal contribution of institutions to the process of knowledge spillovers. We then implement this framework and evaluate the marginal impact on knowledge spillovers played by a specific institution, considering the case of biological resource centers in the life sciences.

III. Biological Resource Centers and their Role in the Life Sciences

III.A. What are Biological Resource Centers?

Biological Resource Centers collect, certify and distribute biological organisms for use in biological research and in the development of commercial products in the pharmaceutical, agricultural and biotechnology industries. As a key element of the life sciences research infrastructure, BRCs maintain a large and varied collection of biological materials, including cell lines, micro-organisms, recombinant DNA material, biological media and reagents, and the information technology tools that allow researchers to access biological materials. Over the past quarter century, they have come

to play an increasingly important role in scientific and commercial research. For example, since the 1980s, select BRCs, such as the American Type Culture Collection (ATCC), have been critical to the extension of intellectual property rights, by serving as International Patent Depositories for all patented living organisms.

At one level, BRCs serve as a library, making the materials and research results developed by one generation of researchers available to future research endeavors. At a slightly more subtle level, BRCs serve to enhance the validity of research itself by providing a transparent and standardized way of accessing biological materials. The value created by the certification and distribution of biological materials arises from the very nature of how biological research is conducted. Biological research depends on the effective development and implementation of careful experiments that allow researchers to disentangle alternative hypotheses about the composition and functioning of living organisms. In many cases, the key to effective experimental design is to understand detailed properties of a biological organism in order to rule out alternative effects and mechanisms. By using biological materials whose properties have been characterized by prior researchers and which can be accessed through a BRC, scientists can dramatically reduce experimental uncertainty -- the uncertainty associated with the scientific tests themselves. As an economic institution, BRCs therefore reduce experimental uncertainty by providing independent access to a wide variety of standardized biological materials.

To see the role of BRCs more clearly, it is useful to compare them with alternatives for collecting, certifying, and circulating biological materials: peer-to-peer networks, private culture collections, and for-profit culture distributors. Peer-to-peer networks consist of informal exchanges among researchers and are dependent on researchers maintaining culture collections within their laboratories and fulfilling

requests for distribution by others in the research community. Private collections, such as those within individual companies or universities, are less idiosyncratic than peer-to-peer networks but remain dispersed and usually offer only minimal certification and assurances of quality. While for-profit culture distribution firms often offer high-quality products, for-profit firms lack appropriate incentives to undertake the full range of collection and certification activities necessary for achieving the highest rate of scientific and technological progress.

The sub-sections below review four key features of biological resource centers that distinguish them as institutions: certification, preservation, independent access, and scale and scope economies.

III.B. Certification in Biological Resource Centers

A key function performed by BRCs is the *certification* of research materials. While BRCs do not fully replicate experiments published in the scientific literature, all materials incorporated into BRC collections undergo a series of reviews and tests to establish the identity and biological viability of the material.[4] BRCs therefore provide the means for scientific replication. Sophisticated BRCs, such as the ATCC, offer a classification system allowing researchers to evaluate the degree of confidence associated with specific deposits.

Though seemingly straightforward, the certification function is critical to effective life sciences research. Consider the early history of peer-to-peer networks. As described by Michael Gold, peer-to-peer networks in the 1960s and 1970s were ineffectively monitored, resulting in the widespread distribution of misidentified cell cultures. Most

dramatically, a significant portion of laboratory cultures in the United States and throughout the world were overtaken by a strain of the HeLa cell line.[5] The consequences of misidentification are far-reaching. Not only does misidentification cast a cloud over the findings of current researchers (with career implications for those whose results are under suspicion), but also confusion and uncertainty places a longer-term cost on progress. Researchers must painstakingly re-establish the validity of specific findings in order to design and implement new research. In short, certification allows researchers to build on the insights of prior research, avoid needless and costly duplication, and so increase research productivity over time.

One of the key consequences of certification is more effective *standardization* of biological models and experimentation procedures. The value of a specific biological material or model tends to increase with its use by prior researchers, since prior use tends to reduce the degree of experimental uncertainty associated with a given investigation. This results in a positive feedback loop, with increasing use of a small number of biological models for an ever greater number of experiments.[6] BRCs create a common database from which to draw materials, documenting the use of materials by other researchers (through the standardized use of accession numbers and the like), and actively monitoring trends in the use of materials within the research community, BRCs may increase the strength and effectiveness of these network benefits and enhance the use of appropriateness of standardized biological materials.

---

[4] The American Type Culture Collection (ATCC), for example, regularly issues statements notifying researchers about errors that the ATCC has identified and cell lines that had been misclassified.
[5] Ironically, the HeLa cell line (named for the donor Helen Lattimer) was the first *in vitro* cell line to be successfully grown within a laboratory and subsequently transported across long distances (Gold, 1986).
[6] This dynamic is similar to the process of standardization and lock-in found in many other high-technology areas, such as computer software and telecommunications equipment. Economists have paid increasing attention to the impact of "network externalities" over the last decade, developing implications for antitrust and intellectual property policy (Shapiro and Varian, 1999).

III.C. Preservation of Biological Materials

A second key function that biological resource centers serve is the *preservation* of biological materials. BRCs collect, characterize and maintain a richer and more varied collection of biological materials, particularly those whose value is not initially understood, than alternative organizational forms. For example, Kary Mullis' ability to develop the extremely influential PCR technique in the late 1980s relied heavily on the fact that the ATCC had maintained long-term storage on a strain of extremophiles, *Thermus aquaticus*, whose value could not have been predicted at the time or until many years after initial discovery.

On the one hand, the dispersed nature of the peer-to-peer network results in a tremendous amount of replication with little incentive for any one laboratory to maintain the full range of materials of potential use by researchers at other laboratories.

More importantly, the maintenance of materials in the peer-to-peer network depends on specific individuals, raising the possibility that materials will be lost due to retirement or inattention by culture curators. For example, in early 2002, three private university collections have been identified as "orphans" available for new storage site; two of these three were classified as "defunct" by July, 2002.[7]

At the same time, the intellectual property held by for-profit laboratories exists for only a modest time (often less than the time between initial characterization and greatest potential use) leaving the for-profit community with few incentives to indefinitely maintain the widest range of materials. Indeed, for-profit distributors of biological materials have tended to succumb to "cherry-picking," focusing on a narrow range of

materials offering high margins and low storage costs. Because for-profit firms are less likely to internalize the full value of long-term variety, non-profit BRCs play a special and critical role in the development of an effective method for collecting, characterizing, and distributing biological materials.

Moreover, BRCs preserve a permanent record of the flow of biological materials across researchers and time. By formalizing and documenting the exchange and use of biological materials, BRCs play a critical role in the management of biological knowledge. For example, the use of BRC materials allows for the rapid assessment of the novelty of claims made in scientific research papers and patent applications. By reducing the costs associated with the assessment of claims, BRCs may to enhance the productivity of research activities.[8]

III.D. Independent Access provided by Biological Resource Centers

Third, because BRC materials are equally accessible to all members of the scientific and technological community, BRCs encourage *independent access* to the results of prior scientific research. In non-BRC networks, access to source materials is dependent on "good will" of researchers who maintain active cell cultures within their laboratory; such goodwill is difficult to maintain when researchers are simultaneously competing with each other to establish new research findings or when the goal of a particular experiment may cast prior findings in an unfavorable light. Alternatively, for-profit characterization and distribution companies will often find it in their private interest (though not in the social interest) to arrange for exclusive access to their

---

[7] *See* http://methanogens.pdx.edu/usfcc.

[8] In some circumstances, this documentation serves as a critical national security resource. For example, the recent anthrax investigations have been impeded substantially by the absence of a centralized database of exchanges of biological materials; relative to peer-to-peer exchanges or even for-profit laboratories, BRCs are recognized for their ability to systematically track the flow of biological materials over time.

databases and materials; recent controversies over the "ownership" of the results of the Human Genome Project are but the most visible in the ongoing war over access to biological materials and data.

III.E. Scale and Scope Economies

Finally, as "living libraries" that continuously collect material developed by the scientific community, BRCs are able to achieve substantial scale and scope economies. Relative to other organizational forms that preserve life science materials, BRCs maintain larger, more varied, and more balanced collections. As a result, BRCs are more likely to undertake the investments that are necessary to increase the quality and reduce the cost of accessing biological materials. For example, over the past decade, institutions such as the ATCC, the Coriell Institute, and the Jackson Laboratory have each established a position of global leadership in specific materials and collection areas. This scale has coincided with a substantial commitment to high quality levels for each activity under its domain. As a consequence of these investments, these BRCs are able to offer access to a larger, more diverse, and more balanced collection at a lower cost than alternatives. These scale and scope economies are reflected in the increasing use of non-profit BRCs for private collections (e.g., by private pharmaceutical and biotechnology companies) and in the successful implementation of BRCs as official international patent depositories. In contrast, in the more dispersed peer-to-peer network, duplication abounds across laboratories and there are few incentives to maintain the high quality levels or the broadest portfolio. Another advantage of a broad portfolio is the accession of materials whose initial value is uncertain; a wider collection allows the life sciences community to maintain an "option" on biological materials. Particularly in the evolving bioinformatics

era, exploiting scale and scope economies through BRCs is crucial for the increased intensity of materials use by life sciences researchers.

IV. An empirical framework for assessing the impact of institutions on knowledge spillovers

This section outlines an empirical framework that allows us to evaluate the impact of BRCs on spillovers of knowledge. We assess the impact of BRCs by exploiting the fact that their impact is made visible through the pattern of article citations in the scientific literature. Citations provide a useful (though noisy) index of the "impact" of an academic article on subsequent scientific research. If depositing biological materials in BRCs is an important ingredient in the process of cumulative research, then scientific articles associated with BRC deposits should be more intensively cited as a result of their greater impact on follow-on research.

The principal issue that our framework addresses is the challenge of isolating the empirical impact of specific institutions, such as BRCs, on the cumulativeness of knowledge production. Specifically, it is difficult to disentangle the role played by institutions from the characteristics of a piece of knowledge that cause it to have an impact (e.g., its intrinsic importance or the network position the initial discoverer). We address this challenge by exploiting the subtle institutional variation in BRC deposits to evaluate a "differences-in-differences" estimator. To do this, we exploit the fact that various subsets of BRC deposits have been shifted exogenously from prior institutional arrangements into biological resource centers. For example, some collections that are maintained in a private university laboratory may be shifted into a public BRC if the principal investigator retires or switching universities.

By comparing citation patterns between a sample of articles linked to BRC deposits with those of a control group (chosen as the preceding articles in the same issue of the same journal), we can ascertain whether knowledge associated with BRC materials has a greater than average impact on future research. This result may obtain, however, simply because researchers deposit materials that are intrinsically important. To distinguish this 'selection effect' from the marginal impact of the BRC on knowledge spillovers, we exploit the experiments associated with a few instances in which collections of materials were shifted exogenously into biological resource centers. By evaluating whether articles associated with such materials receive a 'boost' in citations (relative to the within-article trend, controlling for the age of the article as well as time period effects), we obtain an estimate of the marginal impact of the BRC on knowledge cumulation.

V. Data

V.A. Dataset construction

To create a dataset that allows us to apply the empirical methodology described in the previous section, we overcome two main challenges. First, while most prior use of citation data focuses on the affiliations of the authors of the research, here we are interested in identifying a set of research articles associated with BRC deposits. Second, we need to design a database that allows us to identify the effects associated with selection into at BRC as well as the marginal impact of BRCs on knowledge diffusion.

In order to build a sample of research articles associated with BRC deposits, we take advantage of the fact that ATCC prepares reference information material deposited in its collection. For each material available from ATCC, this information records the

name of the original depositor, the date of the deposit, and key scientific articles associated with the deposit. The ATCC catalog (maintained online at www.ATCC.org, and historically published in catalog-form) identifies the references associated with ATCC deposits, as well as other information on the material. For each deposit, we consider the first article listed within the ATCC deposit reference section as the "focal" article associated with that deposit.[9]

We use this information to construct a dataset comprised of two major sub-samples. The first sub-sample, which we refer to as "the base sample," includes a random selection of articles associated with materials deposited at ATCC by various researchers. The second sub-sample, which we refer as "the special collections" sample includes articles associated with particular "special collections" that were transferred in bulk to ATCC from private culture collections.[10]

To compile the base sample, we have randomly selected a set of 190 deposits from among the materials deposited in three of ATCC's primary collections (Bacteria, Cell Biology, and Molecular Biology).[11] Since these articles have not been exogenously assigned to ATCC, it is not possibly to empirically separate the intrinsic value of the articles from the increment to their value that accrues as a result of their deposit. By comparing the citation pattern of the articles in this sub-sample with those of suitable controls, we can identify whether ATCC articles achieve greater diffusion than average scientific articles.

---

[9] Multiple members of the scientific and information technology staff at ATCC with whom we conducted interviews suggest that the first reference article is typically the one most closely associated with initial use of the biological material.

[10] Numerous scientists, research institutions, and corporations maintain private collections. With the exception of those collections operated by firms, many of these allow open access to their collections; on balance, however, they are less engaged in characterization and knowledge of the contents of their collections is less well-diffused.

[11] Deposit dates for these materials ranged from 1984 to 1999.

In order to identify the impact ATCC-affiliation on the articles in the dataset, each is matched with an associated control article. To ensure that the control article is similar to the ATCC-associated article on as many observable dimensions as possible, we select as a control the article that immediate precedes the focal article in the journal in which it was published. For example, if an ATCC-associated reference were the third article in the June 14, 1986 issue of *Cell*, our control article would be the second article within that same issue.[12] As a result of our choosing the treatment and control articles in this way, both the BRC-affiliated article and the control article will have undergone the same scientific review process and been published at the same moment in time. Consequently, comparing the patterns of citations by future researchers to these articles provides an indication of the relative impact of the two articles on subsequent scientific research. We identify control articles via the PUBMED database of scientific journals. We compile additional article-specific data and tabulate article annual citations from the Institute of Scientific Information's database the Science Citation Index (SCI). By comparing the citation patterns of articles in the base sub-sample with the citation patterns of their control articles we can measure to which knowledge associated with ATCC disseminates in the scientific community.

The data in second sub-sample in our dataset allows us to separately identify both the selection effect and the marginal impact of ATCC-deposit on knowledge diffusion. To do this, we take advantage of the fact that some materials available from ATCC have been transferred in bulk from other collections. Such exogenous transfers occur, for example, when scientists who have maintained private collections retire or when

[12] In the event that the ATCC-associated article is the lead article in its particular issue, we use the second article in that journal as the control.

university funding exigencies necessitate a collection's being moved from its original home.

Of the ATCC special collections, there are four whose accession into ATCC appears to be particularly exogenous.[13] The articles associated with these deposits constitute "special collections" data. The first special collection is a set of articles associated with the Gazdar Collection. This collection was transferred into the ATCC when Dr. Adi Gazdar left his position as Head of Tumor Cell Biology Section at the National Cancer Institutes, along with his collaborator, Dr. John Minna, to become Professor of Pathology at the Hamon center for Therapeutic Oncology at UT Southwestern. The materials in the Gazdar collection were accessioned beginning in 1994. The second set of materials is drawn from the Tumor Immunology Bank (TIB), which was transferred from the Salk Institute in 1981 due to funding considerations and was accessioned beginning in 1982. The third set of articles in the dataset is associated with materials in the Human Tumor Bank (HTB). The HTB had been maintained by researchers at Sloan-Kettering until funding considerations led to its being transferred into ATCC beginning in 1981. The final set of articles in the special collections sub-sample includes a set of articles associated with the David Nanney/Ellen Simon Protistology Collection, which was accessioned into ATCC based on a private endowment from Dr. Ellen Simon.

The special collections sub-sample consists of six articles associated with the Gazdar Collection, 77 with the TIB Collection, 44 with the HTB Collection, and ten with the Protistology Collection. We match each of these articles with a control article in the same manner as that described for the base sample. This structure allows us to construct

a differences-in-differences specification to identify both the selection effect as well as the marginal effect of ATCC deposit on the citation trajectory of those articles. The selection effect is apparent in the differences in the citation patterns of ATCC vs. control articles, controlling for the marginal impact of accession. The marginal impact of ATCC deposit is then evident in the change in the citation trajectory that occurs *after* the special collection is accessioned into ATCC, controlling for all other factors (including year, article vintage, and other article characteristics).

V.B. Summary Statistics

For the variables used in our analysis, Table 1 provides variable names and definitions, while Table 2 reports summary statistics. Our complete dataset consists of the base sub-sample, the special collections sub-sample, and their associated control articles. For each of the articles in the dataset, we track citations beginning in the year in which the article was published and continuing until 2001. The total number of articles in the dataset is 640. and the total number of article-year observations is 10,542. The overall distribution of "vintages" from which we draw article is displayed in Figure A.

The key dependent variable in our analysis is FORWARD CITATIONS, the number of articles that reference the focal article in a given year. In the overall sample, the average level of citation is quite high, relative to traditional measures. In part, this is because the science associated with BRC deposits (and the control articles) tends to be in top-tier journals (e.g., *Science, Nature,* and *Cell).* As well, and consistent with most citation analysis, the distribution is quite skewed (Figure B). As of the end of 2001, the average number of total citations is nearly 70. The average annual FORWARD

[13] Historical details on ATCC's collections are drawn from discussions with Dr. Robert Hay, director of the

CITATIONS varies greatly across collections (Table 3). The articles associated with the Gazdar collection receive more than 22 citations per year, while the HTB and TIB collections receive approximately 11.5 citations, and the Protistology collection receives approximately 1 citation annually. In the base sample and each of the special collection samples, FORWARD CITATIONS to ATCC articles substantially exceeds FORWARD CITATIONS to control articles: FORWARD CITATIONS to ATCC articles in the base and protistology samples are nearly 100% greater than to associated control articles; the difference is more than 800% for the Gazdar collection.

Because the dataset, by construction, contains an equal number of ATCC and non-ATCC articles, the mean of ATCC ARTICLE equals 0.5. The key control variables are the calendar YEAR (ranging from 1970 through 2001) and the VINTAGE, the number of years since the article's initial publication. For each article, we also record a PUBLICATION YEAR; for articles in the special collections we also include a DEPOSIT YEAR, which reflects the year in which the material associated with that article was accessioned into the ATCC collection.[14] The apparent oddity that the average PUBLICATION YEAR is greater than the DEPOSIT YEAR is explained by the fact that PUBLICATION YEAR includes articles in the base sample, which have been chosen randomly from among all articles associated with ATCC materials between 1970 and 2001, while the articles for which DEPOSIT YEARs are recorded include only those in the special collections, none of which were accessioned prior to 1982. For each of the

---

Department of Cell Biology at ATCC.

[14] In some cases, the DEPOSIT YEAR is measured with error (of up to a few months). As materials moved wholesale into ATCC must undergo authentication and cataloging before they are available to public use, there is some delay between the announcement of a transfer and ATCC's ability to ship materials for scientific use. In some occasions, materials may be available for a few months before their accession is officially declared in a catalog or other ATCC publication.

materials in the special collections, we also track the current PRICE; this averages approximately $225 per order.

While our current analysis focuses mostly on specifications that address article heterogeneity by including article fixed effects, we have collected systematic characteristics about each of the articles in our sample. Specifically, we have information on the number of pages for each article (# PAGES), the number of authors (# AUTHORS), the number of backward citations (BACKWARD CITATIONS). In addition, we record whether the lead author is associated with a university (UNIVERSITY), government (GOVERNMENT), and whether their address is foreign or domestic (NON-US). The lead authors of majority of articles in the sample are affiliated with a university (59%); 14% are affiliated with a government agency; and 33% are not from the United States.

VI. Empirical Results

Our empirical work is divided into two parts. In our baseline analysis, we employ data from the base sample in order to compare citation patterns of ATCC articles with non-ATCC articles (Table 4). These results demonstrate that ATCC articles are more highly cited than controls. This methodology, however, cannot isolate selection effects from the marginal-ATCC effects. The special collections data allow us to separately identify the importance of these effects based on a nuanced differences-in-differences analysis (Tables 5-7). This analysis relies on variation arising from a change in the status of whether an article is associated with a BRC deposit. By simultaneously comparing citation patterns *across article pairs* (i.e., comparing articles eventually deposited in BRCs with those that are not) and *across deposit-status within article* (i.e., whether a

particular article has yet been deposited), we can identify the selection effect separately from the marginal impact of ATCC deposit. A positive, significant fixed effect for articles that eventually get deposited at ATCC (ATCC ARTICLE) implies a selection effect (controlling for all other factors). Evidence of a marginal impact of ATCC on knowledge diffusion would arise if a boost (or decline) in FORWARD CITATION occurs subsequent to deposit with ATCC (ATCC ARTICLE, POST DEPOSIT), controlling for all other factors, including whether the article is ever deposited with ATCC. The results demonstrate economically significant effects of both selection and ATCC-impact.

VI.A. Baseline analysis

Table 4 begins with a straightforward OLS specification of LOG CITATIONS on ATCC-ARTICLE, including fixed effects for each article "pair," vintage year, and calendar year. Thus, (4-1) evaluates the difference in citations between ATCC-linked and non-ATCC-linked articles, controlling for the article pair, the year in which the articles were published, and the amount of time that has passed since the publication (i.e., the articles' "vintage"). The results evidence a significant impact of ATCC-association: on average, ATCC-referenced articles receive 72% more citations per year than non-ATCC articles. Figures C and D portray this striking disparity. Figure C graphs the distribution of differences in citations between ATCC-linked articles and controls; Figure D graphs these differences by article vintage year (D-1 presents the differences in levels, D-2 in percentages). While both ATCC-referenced and control articles are highly cited, ATCC articles consistently have a higher rate of citation. Moreover, the "citation

premium" received by ATCC articles tends to increase, as a percentage of citations, over the first twenty years after an article's publication.

Of course, the use of LOG CITATIONS or a simple unconditional graph is problematic because citation data are highly skewed. In this circumstance, count data methods are more appropriate. In (4-2), we turn to a negative binomial regression (a Poisson approach which relaxes the equality of mean and variance), using the same variables as in (4-1). The coefficients in these models are reported as incidence-rate ratios. (Thus, coefficients equal to one imply no effect on FORWARD CITATIONS; a coefficient equal to 1.50 implies a 50% boost to FORWARD CITATIONS.) After accounting for the skewness of the data, the results evidence an even stronger quantitative impact of ATCC association. In each of the cross-sectional binomial regressions, we find that ATCC-referencing articles receive more than twice as many citations per year than control articles. (4-3) demonstrates that the effect of ATCC affiliation is positive for each of the special collections, although there are differences across collection.

In addition to the results in Table 4, we perform a specification which includes fixed article effects, fixed vintage effects, and fixed calendar year effects. The vintage effects and article fixed effects are highly significant, consistent with the fact that citation patterns are highly skewed and most scientific publications have a well-defined "lifetime" of impact. Figure E graphs the estimated conditional vintage effects, while Figure F maps the overall distribution of article-specific effects. Each of these calculations is computed while taking the other sources of heterogeneity into account and therefore provides a more nuanced picture of the "true" impact of vintage on performance.

VI.B. Separately identifying selection effects and the marginal impact of ATCC deposit

Motivated by these statements about the strong impact of heterogeneity on the data, we now turn to our differences-in-differences analysis in Table 5. Equation (5-1) precisely identifies the differential effect of selection versus ATCC-impact by including an indicator variable for ATCC-referencing articles (ATCC ARTICLE) as well as a separate variable that identifies ATCC articles *after* they have been deposited (ATCC ARTICLE, POST DEPOSIT) in a model that also includes controls for vintage effects, year effects, and article pair effects. Conditional on the article pair, the incidence rate ratio on ATCC ARTICLE implies that ATCC-referencing articles receive 181% more citations than control articles. In the same model, the incidence rate ratio on ATCC-ARTICLE, POST DEPOSIT indicates that in the years subsequent to their deposit ATCC-referencing articles receive an additional 99% boost in their citation frequency. While the magnitude of the selection effect substantially exceeds the post-deposit impact of ATCC association on the citation frequency of ATCC-referencing articles, the economic importance of each is remarkable.

The remainder of the models in Table 5 focus on the post-deposit impact of ATCC association. Each includes article fixed effects, which absorb the effect of selection into ATCC. Equations (5-2) and (5-3) demonstrate that neither including article fixed effects nor correcting for CUMULATIVE CITATIONS obviates the impact of ATCC-deposit found in (5-1). By including a term representing the interaction between ATCC-ARTICLE*TIME, (5-4) demonstrates that the post-deposit impact of ATCC association even grows, on average, over time.

VI.C. Additional Examinations on the Special Collections

These results are robust to a number of alternative specifications, sample definitions, and econometric treatments. For example, in Table 6, we conduct separate analyses for each of the four special collections, allowing separate calendar and vintage effects for each of the four samples and controlling for article fixed effects (thus running the equivalent of (5-2) for each sample. With the exception of the Protistology Collection, the post-deposit impact of ATCC association is positive, statistically significant, and economically important. The Gazdar and HTB collections evidence a 52% post-deposit boost in citation frequency, while the TIB collection receives a 105% boost. Unlike the other collections, the Protistology Collection, which, like the Gazdar Collection, was deposited in the 1990s and therefore has a relatively smaller number of observations obtains neither a positive nor statistically significant citation boost from ATCC deposit. These results suggest that the post-deposit impact of ATCC association does vary slightly by collection, although it is greater than 50% in most cases.

For Figure G we run collection-specific specifications similar to (5-2), which include specific dummy variables for each year prior to and since deposit. Figure G plots the results, focusing on the years immediately prior to and subsequent to deposit. Consistent with earlier analyses, average post-deposit citation frequency is substantially greater than pre-deposit citation frequency across the collections. Further, the analysis demonstrates that the impact of ATCC deposit increases markedly over time. The value of deposit rises slowly at first, but increases substantially over time.

The pattern of citations in the few years prior to accession deserves further attention. These years correspond to the time during which the special collections are about to be moved to ATCC, but have not yet officially entered the ATCC collection.

The "accession date" is necessarily measured with some error, particularly those of the HTB & TIB collections, which were accessioned in the early 1980s. ATCC data do indicate when a cell line became officially available; however, public announcements were not made every time a particular cell line became available for delivery. While Figure G demonstrates no important discernible upwards trend in the pattern of pre-deposit citations associated with the Gazdar, HTB, and Protistology collections, citations to the TIB collection do trend upwards in each of the four years prior to deposit. This calls into question our certainty in the exogeneity of the deposit of the TIB collection. We therefore omit this collection from each of our further analyses. We also omit the Protistology Collection, for which (6-4) identified no significant post-deposit impact on citation.

VI.D. Analysis of Robustness

Table 7 explores the robustness of the results to the omission of the TIB collection, as well as to addition modifications. Equation (7-1) re-estimates (5-2) without the TIB and Protistology collections. The impact of these omissions is slight. The impact of ATCC deposit on citation frequency remains significant and greater than 40%. Thus, even in our most conservative estimate of the impact of ATCC deposit on citation patterns, exogenously articles deposited with ATCC receive more than 40 percent greater citations than they had prior to their deposit.

Clustering standard errors by article rather than article pair, equation (7-2) demonstrates that the result are not sensitive our specification of the structure of the standard errors.

VI.E. The impact of deposit over time

We consider how the impact of BRC deposit has changed over calendar time in Figure H. This graph reports the coefficients from a regression similar to (5-2) that includes dummy variables for the interaction between each calendar year since 1984 and POST-ATCC-DEPOSIT. The results are quite intriguing. While the value of BRC deposit was insignificant in the mid-1980s, the returns have steadily increased over time (they become consistently significant in a statistical sense after 1989). Perhaps more pointedly, there seems to have been an acceleration after 1990.

VI.F. Assessing the Cost-Effectiveness of BRCs

In the final step of our empirical analysis, we review the cost-effectiveness of biological resource centers. A comprehensive cost-benefit analysis is beyond the scope of our analysis, particularly because we cannot fully capture the degree to which access to BRC materials improves research productivity of users. We can, however, assesses the extent to which a given level of expenditures on BRC deposit compares to alternative research in promoting cumulative progress. Specifically, our analysis compares investments in BRC deposit and authentication activities to traditional grant programs with respect to their efficiency in seeding the knowledge stock of future researchers. Our exercise involves the calculation of three estimates:

The first step is obtaining a baseline citation cost, i.e., the "cost per citation" paid by public funding agencies (such as NIH) when allocating resources that result in published scientific articles. This estimate is calculated using the estimates in Adams and Griliches (1996). Using data drawn from the 1980s, Adams and Griliches estimate the relationship between expenditures and academic research output (papers and citations)

for individual academic departments at top universities across the United States, including biology departments. Using these measures (and converting all expenditures into 1987 current dollars), they estimate the cost per citation to be $2400 for expenditures at a top-ten biology department and at $4200 for citations at non-elite public universities. Using the BEA R&D price deflator to restate this figure in current dollars, the *lowest* Adams and Griliches estimates of current cost per citation is $2887. Being conservative (in terms of estimating the effectiveness of BRC expenditures), we choose the lowest estimated cost per citation among these figures, and so set the Baseline Citation Cost at $2400 for the life sciences.

The second figure we incorporate in the analysis is the BRC Accession Cost: The full cost of deposit and accession into a national BRC collection such as the ATCC. The recent OECD Report on Biological Resource Centers (2001) provides estimates of this cost from BRCs based on a recent survey; the highest estimate of BRC Accession Cost according to the OECD report is $10,000 (this was the maximum of the range of the survey response given by the ATCC). While it is likely that the true marginal accession cost may be somewhat lower than $10,000, we use this high-end figure to bias us away from finding evidence for cost-effectiveness on the part of BRCs.

Finally, we employ these figures to compute the BRC Citation Boost, equal to the incremental number of citations expected to result from deposit and accession into a national BRC. We compute three different estimates of the BRC Citation Boost (Table 8). The first two of these computations builds on the data provided by Adams and Griliches (1996). In their work, the average biology publication received 24.6 citations during the first five years of publication if authors were located at a top ten university and 14.3 citation if authors were located at universities below the top ten (in biology). As

well, in our most conservative estimate, BRC deposit was associated with an 81% increase in citations. If we assume that the marginal accessioned material comes from a top ten university laboratory, then the marginal impact from deposit is estimated to be 19.9; if the accessioned materials is drawn truly at random, we assign the citation impact to be 11.6, based on the citation rates of articles published by authors outside the top ten. We also compute the BRC Citation Boost directly from the estimates provided in the last section, focusing on the incremental boost realized by BRC-linked articles within the sample. Using this formulation, the BRC Citation Boost is 20.5; interestingly, BRC-linked articles within the sample have a BRC Citation Boost quite close to the estimated BRC Citation Boost for articles which would be drawn from top-tier biology departments themselves.

Dividing the BRC Citation Boost by the BRC Accession Cost yields an estimate of the BRC Citation Cost which we can then compare with the Baseline Citation Cost. These estimates are dramatic. Even imposing the estimates that result in a conservative calculation, BRC deposit expenditures offer nearly a three-fold efficiency benefit in terms of inducing citations. For articles that have been deposited in the ATCC collections, this efficiency boost is estimated to be nearly five-fold. It is important to interpret the calculations cautiously because of the noisiness of citation data. To the extent, however, that the primary criterion for current public basic research expenditures at NIH is the likelihood that such research will have important disciplinary impact (which is often measured through citation counts), this analysis suggests that depositing research materials in biological resource centers may substantially amplify the impact of (or rate-of-return on) already funded and published research.

VII. DISCUSSION

This paper characterizes the impact of institutions on knowledge Cumulation as having two components. The first of these, which we term a selection effect, acknowledges that knowledge associated with a particular institution may spillover in a quantity that covaries positively with the quality of the individuals and concomitant research affiliated with that institution. The second of these, which we describe as the marginal impact of the institution, refers to the incremental impact that an institution has on the contribution of a piece of knowledge to the overall stock of knowledge, conditional on its quality.

The results of our empirical analysis suggest that biological resource centers play a subtle but crucial role in sustaining R&D productivity in scientific and technological disciplines. Knowledge associated with materials deposited in BRCs evidences a substantially greater impact on future research than controls – implying that BRCs serve as repositories for materials that are important to life sciences. In addition, our analysis demonstrates that depositing materials in BRCs significantly *amplifies* the impact of knowledge associated with those materials.

REFERENCES

Ambramowitz, M. (1956). "Resource and Output Trends in the United States since 1870," *American Economic Review* 46: 5-23.

Adams, J. (1990) "Fundamental Stocks of Knowledge and Productivity Growth," *Journal of Political Economy*, 98, 673-702.

Ambramovitz, M. (1956). "Catching Up, Forging Ahead and Falling Behind," *Journal of Economic History*, 46, 385-406.

Branstetter, L. (2000) "Exploring the Link Between Academic Science and Industrial Innovation: The Case of California's Research Universities," mimeo, prepared for the 2000 NBER Summer Institute.

Branstetter, L. and M. Sakakibara (2001) "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms," *Rand Journal of Economics,* 32(10), 77-100.

Bush, V. (1945). *Science: The Endless Frontier*. Washington (DC): United States GPO.

Coe, D. and E. Helpman (1995) "International R&D Spillovers," *European Economic Review*, 39, 859-887.

Dasgupta, P. and P. David (1994) "Towards a new economics of science," *Research Policy*, 23, 487-521.

Eaton, J. and Kortum, S. (1996) "Trade in Ideas: Productivity and Patenting in the OECD," *Journal of International Economics*.

Furman, J., M.E. Porter, and S. Stern (2001). "The Determinants of National Innovative Capacity," *Research Policy*, forthcoming.

Gold., M. (1986). *A Conspiracy of Cells*, State University of New York Press.

Grossman, G. and E. Helpman (1991) *Innovation and Growth in the Global Economy*. Cambridge (MA): MIT Press.

Irwin, D. and P. Klenouw (1996) "High-tech R&D subsidies: Estimating the effects of Sematech," *Journal of International Economics*, 40, 323-344.

Jaffe, A. and J. Lerner (2001) "Reinventing Public R&D: Patent Law and Technology Transfer from Federal Laboratories," *Rand Journal of Economics*, 32, 167-198.

Jaffe, A., M. Trajtenberg, and R. Henderson (1993), Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *The Quarterly Journal of Economics*, 434, 577-598.

Jones, C. (1995). "R&D Based Models of Economic Growth," *Journal of Political Economy*, 103, 739-784.

Keller, W. (2000) "Do Trade Patterns and Technology Flows Affect Productivity Growth?, *World Bank Economic Review*.

Keller, W. (2002) "Geographic Localization of International Technology Diffusion," *American Economic Review*.

Kortum, S. and J. Lerner (1999) "What is Behind the Recent Surge in Patenting?" *Research Policy* 28, 1-22.

Kortum, S. and J. Lerner (2000) "Assessing the Contribution of Venture Capital to Innovation," *Rand Journal of Economics*, 31 (Winter 2000) 674-692.

Lerner, J. and J. Tirole (2002) "Patent Pools: Theory and Evidence," mimeo Harvard Business School.

Mansfield, E. (1995) "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing," *The Review of Economics and* Statistics, 77, 55-65.

Mowery, D. and N. Rosenberg (1989) *Technology and the Pursuit of Economic Growth*. Cambridge, UK: Cambridge University Press.

Mowery, D. and N. Rosenberg (1998) *Paths to Innovation*. Cambridge (UK): Cambridge University Press.

Mowery, D., R. Nelson, B. Sampat, and A. Ziedonis (2001) "The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980," *Research Policy*, 30, 99-119.

Mowery D. and A. Ziedonis (2002) "Academic Patent Quality and Quantity Before and After the Bayh-Dole Act in the United States," *Research Policy*, 31, 399-418.

Nelson, R.R. (1959) "The Simple Economics of Basic Scientific Research," *Journal of Political Economy* 67(3), 297-306.

Nelson, R. R., ed. (1993) *National Innovation Systems: A Comparative Analysis*. New York (NY): Oxford University Press.

Nelson R. R. and R. Rosenberg (1994) "American Universities and Technical Advance in Industry," *Research Policy* 23(3), 323-348.

OECD (2001)*Biological Resource Centers: Underpinning the Future of Life Sciences and Biotechnology*, Paris, FR: Organization for Economic Co-operation and Development, Directorate for Science, Technology, and Industry,

Porter, M.E. and S. Stern (2000) “Measuring the "Ideas" Production Function: Evidence from International Patent Output,” NBER Working Papers 7891, National Bureau of Economic Research.

Powell, W. (1998) “Learning from collaboration: Knowledge and networks in the biotechnology and pharmaceutical industries,” *California Management Review*, 40(3), pp. 228-240.

Romer, P. (1990) “Endogenous Technological Change,” *Journal of Political Economy,* 98, S71-S102.

Rosenberg, Nathan (1963) “Technological Change in the Machine Tool Industry, 1840-1910,” *Journal of Economic History*, 23, 414-443.

Rosenberg, N. (1974) “Science, Invention, and Economic Growth,” *The Economic Journal*, 84(333), 90-108.

Rosenberg, Nathan (1982). *Inside the Black Box: Technology and Economics*. Cambridge (UK) Cambridge University Press.

Rosenberg, N. and R. Nelson (1993) “American Universities and Technical Advance in Industry,” *Research Policy*, 23, 323-348.

Rosenkopf, L. and M. Tushman (1998) “The Coevolution of Community Networks and Technology: Lessons from the Flight Simulation Industry,” *Industrial and Corporate Change*. 7(2), 311-346.

Sakakibara, M. and L. Branstetter (2001) “Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms,” *Rand Journal of Economics*.

Scotchmer, S. (1991) “Standing on the Shoulders of Giants: Cumulative Research and the Patent Law,” *Journal of Economic Perspectives*, 5(1), 29-41.

Sakakibara, M. and L. Branstetter, (2001) “Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms,” *Rand Journal of Economics*, 32(1), 77-100.

Solow, R.M. (1956). “A Contribution to the Theory of Economic Growth,” *Quarterly Journal of Economics,* 70, 65-94.

Solow, Robert M. (1957). “Technical Change and the Aggregate Production Function,” *Review of Economics and Statistics*, 39:312-20.

TABLE 1
VARIABLES & DEFINITIONS

VARIABLE	DEFINITION	SOURCE
CITATION CHARACTERISTICS		
FORWARD CITATIONS$_{jt}$	# of Forward Citations to Article *j* in Year *t*	Science Citation Index (SCI)
CUMULATIVE CITATIONS$_{jt}$	# of FORWARD CITATIONS from publication date to YEAR$_{t-1}$	SCI
YEAR	Year Trend; also Year Dummy Variables	SCI
VINTAGE	Year – year of publication	SCI
ARTICLE CHARACTERISTICS		
ATCC ARTICLE	Dummy variable equal to 1 if Article is associated with a material deposited in the American Type Culture Collection (ATCC)	ATCC
ATCC ARTICLE, POST DEPOSIT	Dummy variable equal to 1 if Article is reference by ATCC deposit and YEAR > DEPOSIT YEAR (i.e., deposit has already occurred)	ATCC
COLLECTION	Dummy variable indicating the collection with which the article is associated (1 = Gazdar Collection; 2 = Tumor Immunology Bank (TIB); 3 = Human Tumor Bank (HTB)) ***Gazdar Collection***: This collection was transferred into the ATCC when Dr. Adi Gazdar left his position as Head of Tumor Cell Biology Section at the National Cancer Institutes, along with his collaborator, Dr. John Minna, to become Professor of Pathology at the Hamon center for Therapeutic Oncology at UT Southwestern. The Gazdar collection was incorporated into ATCC over a number of years; the materials examined in this paper were accessioned into in 1994. ***TIB Collection***: The Tumor Immunology Bank (TIB) was created at ATCC when a collection was transferred from the Salk Institute in 1981, and accessioned into the ATCC over the next few years. ***HTB Collection***: The Human Tumor Bank was maintained at Sloan-Kettering until 1981; it was accessioned into the ATCC collection over the next few years. ***Protistology Collection***: The Protistology Collection was donated to the ATCC by Ellen Simon in 1998.	ATCC
PRICE	For articles associated with ATCC products, the price at which the ATCC product can be purchased; 0 otherwise	ATCC
DEPOSIT YEAR	Year in which the material associated with Article *j* is "accessioned" and available for purchase through the ATCC	ATCC
PUBLICATION YEAR	Year in which Article *j* is published	SCI
BACKWARD CITATIONS	Number of articles cited by Article *j*	SCI
# PAGES	Count of the number of pages in Article *j*	SCI
# AUTHORS	Count of the number of authors of Article *j*	SCI
UNIVERSITY	Dummy variable equal to 1 if lead author is associated with a university; 0 otherwise	SCI; author verification
GOVERNMENT	Dummy variable equal to 1 if lead author is associated with a government-affiliated institution; 0 otherwise	SCI; author verification
NON-US	Dummy variable equal to 1 if lead author is associated with an institution located outside of the United States; 0 otherwise	SCI; author verification
PRIVATE	Dummy variable equal to 1 if lead author is associated with a private institution; 0 otherwise	SCI; author verification

TABLE 2A
MEANS & STANDARD DEVIATIONS

VARIABLE	N	MEAN	STANDARD DEVIATION
CITATION-YEAR CHARACTERISTICS			
FORWARD CITATIONS	10542	6.11	13.39
CUMULATIVE CITATIONS	10542	67.21	141.80
YEAR	10542	1991.94	6.55
VINTAGE	10542	9.06	6.55
ARTICLE CHARACTERISTICS *(N=640 total articles)*			
TOTAL CITATIONS	640	100.24	188.44
PUBLICATION YEAR	640	1985.53	6.61
ATCC ARTICLE	640	0.50	0.50
DEPOSIT YEAR*	137	1984.54	5.19
PRICE*	137	224.71	46.11
# PAGES	640	7.45	6.04
# AUTHORS	640	3.97	2.54
BACKWARD CITATIONS	640	30.16	23.42
UNIVERSITY	611	0.59	0.49
GOVERNMENT	611	0.14	0.35
NON-US	591	0.33	0.47

* These data exist only for those articles associated with deposits to ATCC; price data are included only for those in the special collections (i.e., the Gazdar, TIB, or HTB collections).

TABLE 3
MEANS & STANDARD DEVIATIONS,
BY COLLECTION & CONTROL GROUP

	COLLECTION									
	Base Sample		Gazdar Sample		HTB Sample		TIB Sample		Protistology	
	ATCC Deposits	Controls	Gazdar Deposits	Controls	HTB Deposits	Controls	TIB Deposits	Controls	Protist. Deposits	Controls
#PAPERS	183	183	6	6	44	44	77	77	10	10
PAPER-YEARS	2429	2429	87	87	854	854	1734	1734	143	143
FORWARD CITATIONS	6.75 (11.03)	3.52 (6.64)	22.28 (33.36)	2.66 (3.96)	11.57 (20.39)	2.32 (6.60)	11.49 (20.82)	2.88 (6.89)	0.98 (1.36)	0.53 (0.72)
CUMULATIVE CITATIONS	89.66 (116.28)	46.77 (67.38)	323.00 (384.76)	38.50 (30.66)	224.59 (299.36)	44.89 (76.84)	260.89 (360.28)	65.46 (110.53)	14.00 (10.30)	7.6 (5.74)
PUBLICATION YEAR	1988.73 (4.68)	1988.73 (4.68)	1987.50 (3.39)	1987.50 (3.39)	1982.16 (6.87)	1982.16 (6.87)	1979.43 (1.85)	1979.43 (1.85)	1988.70 (6.52)	1988.70 (6.52)
DEPOSIT YEAR*			1994.00 (0.00)		1983.14 (2.06)		1982.60 (2.28)		1997.07 (0.57)	
PRICE*			201.30 (32.60)		207.14 (39.06)		244.74 (40.40)		160.00 (0.00)	

* PRICE & DEPOSIT YEAR only meaningful for ATCC deposits (not for Controls)

TABLE 4
CROSS-SECTIONAL RESULTS

	OLS Dep Var = ln(FORWARD CITATIONS)			NEGATIVE BINOMIAL (Coefficients reported as incidence-rate ratios) Dep Var = FORWARD CITATIONS					
	(4-1) Overall ATCC Effect*			(4-2) Baseline Count Model*			(4-3) Auxiliary Count Model^		
ARTICLE CHARACTERISTICS									
ATCC-ARTICLE	**0.72 (0.02)**								
ATCC-ARTICLE, POST-DEPOSIT				**3.08 (0.07)**			**2.19 (0.26)**		
GAZDAR COLLECTION							**5.66 (1.16)**		
HTB COLLECTION							**9.30 (3.72)**		
TIB COLLECTION							**6.76 (1.46)**		
PROTISTOLOGY COLLECTION							**6.90 (1.55)**		
PRICE							**1.00 (0.001)**		
CONTROL VARIABLES									
Parametric Restrictions	# Restrict	F-stat	p-value	# Restrict	χ^2	p-value	# Restrict	χ^2	p-value
Article Pair FEs = 0	**319**	**35.61**	**0.000**	**319**	**106115.1**	**0.000**			
Vintage FEs = 0	**31**	**37.43**	**0.000**	**31**	**805.80**	**0.000**	**31**	**827.05**	**0.000**
Year FEs = 0~	**23**	**12.17**	**0.000**	**23**	**218.88**	**0.000**	**23**	**208.96**	**0.000**
Regression Statistics									
R-squared	0.58								
Log-likelihood				-22906.30			-26260.23		
P-value of Chi				0.00			0.00		
# of Observations	10494			10494			10494		

* Robust standard errors are in parentheses.
^ Robust standard errors, adjusted for clustering by article, are in parentheses.
~ Year FEs included for 1980-2001; 1970-1974 and 1975-1979 grouped.

TABLE 5
TIME-SERIES RESULTS ON "EXPERIMENTAL DATA" ONLY*

	NEGATIVE BINOMIAL (Coefficients reported as incidence-rate ratios) Dep Var = FORWARD CITATIONS			
	(5-1) Selection vs. Shift Effect	**(5-2)** Marginal Impact, with Article FEs	**(5-3)** With Cumulative Citations	**(5-4)** Interactions with Time
ARTICLE CHARACTERISTICS				
ATCC ARTICLE	**2.81 (0.14)**			
ATCC-ARTICLE, POST-DEPOSIT	**1.99 (0.11)**	**1.81 (0.10)**	**1.68 (0.09)**	**1.61 (0.09)**
ATCC-ARTICLE* TIME				**1.05 (0.01)**
CITATION CHARACTERISTICS				
CUMULATIVE CITATIONS			**1.00 (0.00)**	
CONTROL VARIABLES				
Parametric Restrictions	#Restrict χ^2 p-value	#Restrict χ^2 p-value	#Restrict χ^2 p-value	#Restrict χ^2 p-value
Article Pair FEs =0	**136 37659.7 0.000**			
Article FEs =0		**273 67589.3 0.000**	**273 66440.9 0.000**	**273 63355.7 0.000**
Vintage FEs =0	**30 441.54 0.000**	**30 491.69 0.000**	**30 565.91 0.000**	**30 468.70 0.000**
Year FEs =0	**23 54.78 0.000**	**23 93.71 0.000**	**23 119.94 0.000**	**23 138.77 0.000**
Regression Statistics				
Pseudo R-squared	0.19	0.26	0.27	0.27
Log-likelihood	-12355.246	-11219.31	-11138.18	-11186.18
P-value of Chi	0.00	0.00	0.00	0.00
# of Observations	5636	5636	5636	5636

* Robust standard errors are in parentheses.
~ Year FEs included for 1980-2001; 1970-1974 and 1975-1979 grouped.

TABLE 6
TIME-SERIES RESULTS BY COLLECTION*

	NEGATIVE BINOMIAL REGRESSIONS (Coefficients reported as incidence-rate ratios) Dep Var = FORWARD CITATIONS											
	(6-1) Gazdar Collection			**(6-2)** HTB Collection			**(6-3)** TIB Collection			**(6-4)** Protistology Collection		
ARTICLE CHARACTERISTICS												
ATCC-ARTICLE, POST-DEPOSIT	**1.52 (0.31)**			**1.52 (0.14)**			**2.05 (0.14)**			0.96 (0.303)		
CONTROL VARIABLES												
Parametric Restrictions	# Re-strict	χ^2	p-value	# Re-strict	χ^2	p-value	# Re-strict	χ^2	p-value	# Re-strict	χ^2	p-value
Article Pair FEs = 0	**11**	**1197.77**	**0.000**	**87**	**31738.2**	**0.000**	**153**	**31899.8**	**0.000**	**19**	**143.86**	**0.000**
Vintage FEs = 0	**17**	**160.9**	**0.000**	**30**	**633.31**	**0.005**	**30**	**253.87**	**0.000**	**20**	**1797.62**	**0.000**
Year FEs = 0	**17**	**552.9**	**0.000**	**23**	**94.45**	**0.000**	**23**	**49.11**	**0.001**	**21**	**1117.34**	**0.000**
Regression Statistics												
Pseudo R-squared	0.353			0.301			0.256			0.250		
Log-likelihood	-372.63			-3292.53			-7080.29			-257.06		
P-value of Chi	0.00			0.00			0.00			0.00		
# of Observations	174			1706			3470			286		

* Robust standard errors are in parentheses.

TABLE 7
EXPLORING ROBUSTNESS

	NEGATIVE BINOMIAL REGRESSIONS (Coefficients reported as incidence-rate ratios) Dep Var = FORWARD CITATIONS					
	(7-1) (5-2), omitting TIB & Protistology *			**(7-2)** Errors clustered by article^		
ARTICLE CHARACTERISTICS						
ATCC-ARTICLE				**3.99 (0.63)**		
ATCC-ARTICLE, POST-DEPOSIT	**1.43 (0.11)**			**1.36 (0.23)**		
CONTROL VARIABLES						
Parametric Restrictions	# Restrict	χ^2	p-value	# Restrict	χ^2	p-value
Article Pair FEs = 0				**49**	**3030.57**	**0.000**
Article FEs = 0	**99**	**35493.32**	**0.000**			
Vintage FEs = 0	**30**	**742.25**	**0.000**	**30**	**7964.47**	**0.000**
Year FEs = 0	**23**	**132.94**	**0.000**	**23**	**96.78**	**0.000**
Regression Statistics						
Pseudo R-squared	0.301					
Log-likelihood	-3705.49			-4210.24		
# of Observations	1880			1880		

* Note that (7-2) also omits the TIB & Protistology collections.
^ Robust standard errors, adjusted for clustering by article, are in parentheses.

TABLE 8
BRC DEPOSIT COST-EFFECTIVENESS ANALYSIS

Calculation	Baseline Citation Cost	BRC Accession Cost	BRC Citation Boost	BRC Citation Cost	BRC Cost-Effectiveness Index*
"Top Ten" Citation Boost	$2,400	$10,000	19.9	$502	4.78
Random Citation Boost	$2,400	$10,000	11.6	$862	2.78
BRC-Linked Citation Boost	$2,400	$10,000	20.5	$488	4.91

***** BRC Cost-Effectiveness Index = (Baseline Citation Cost)/(BRC Citation Cost)

FIGURE A
NUMBER OF PUBLICATIONS BY YEAR

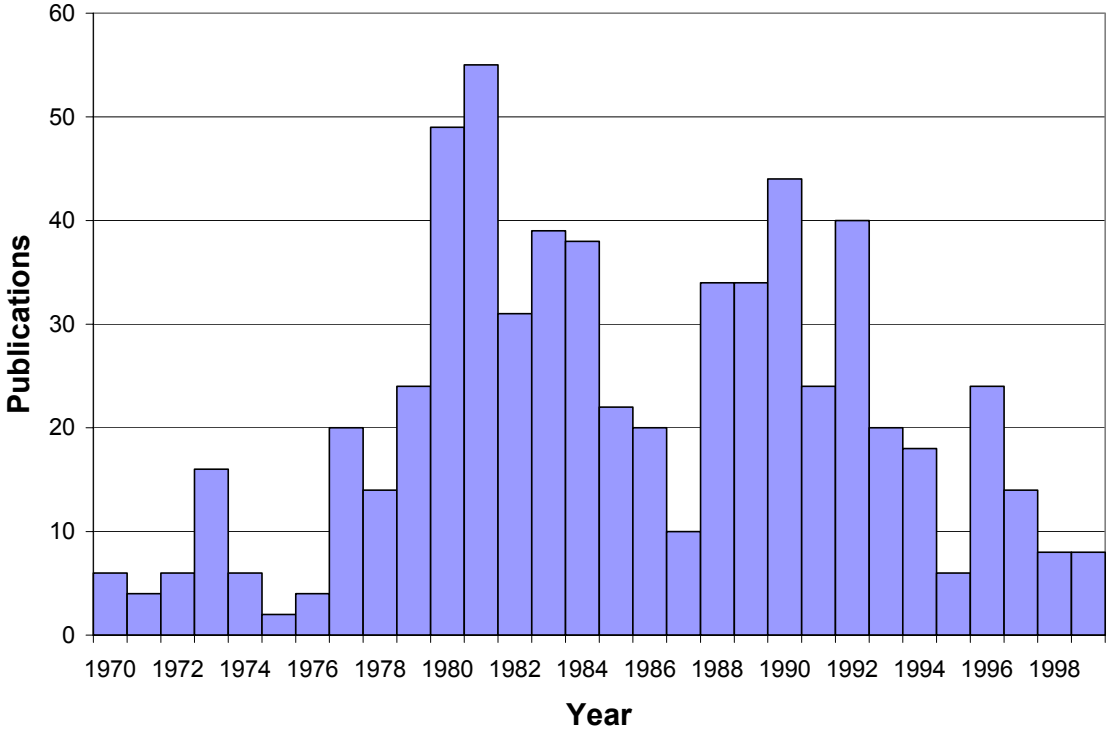


FIGURE B
DISTRIBUTION OF CITATIONS

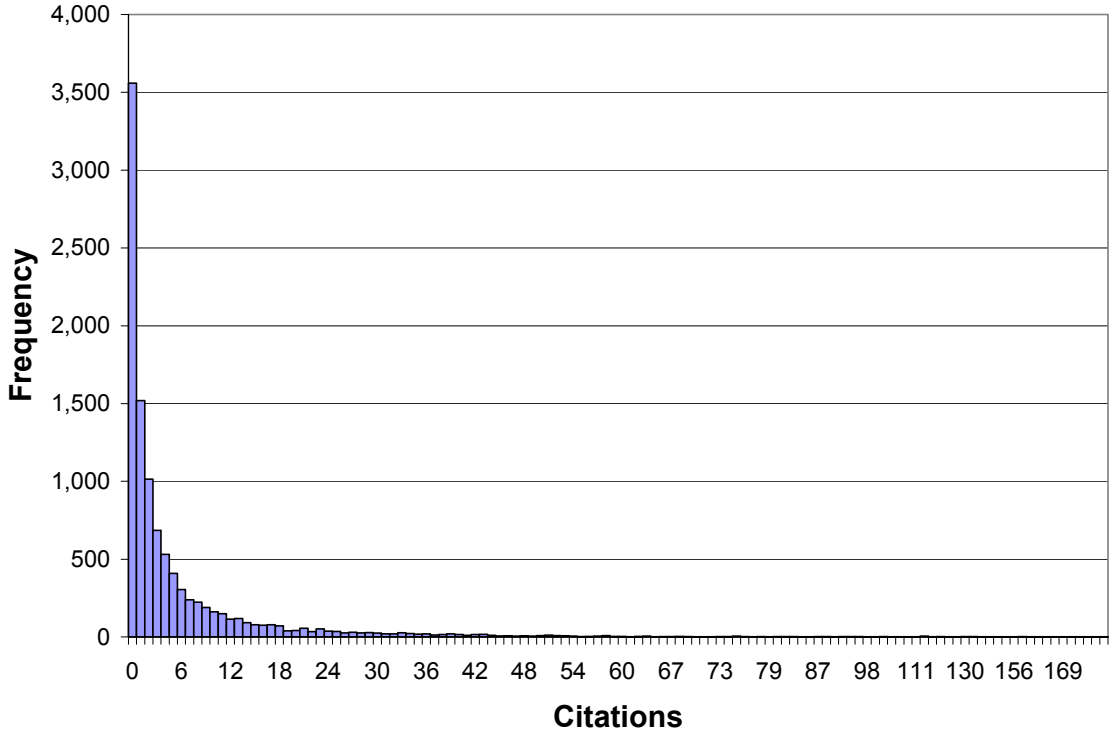


FIGURE C-1
AVERAGE ANNUAL CITATIONS BY VINTAGE,
ATCC VS. CONTROL ARTICLES

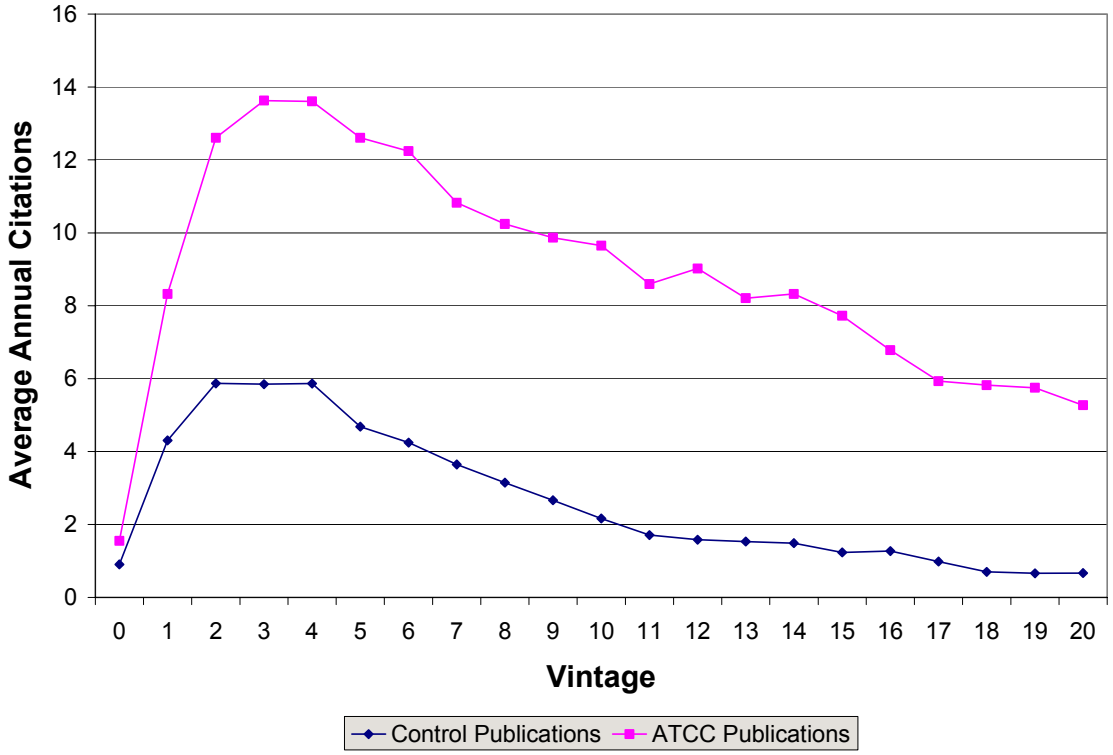


FIGURE C-2
PERCENT DIFFERENCE IN ANNUAL AVERAGE CITATIONS TO ATCC VS. CONTROL ARTICLES, BY VINTAGE

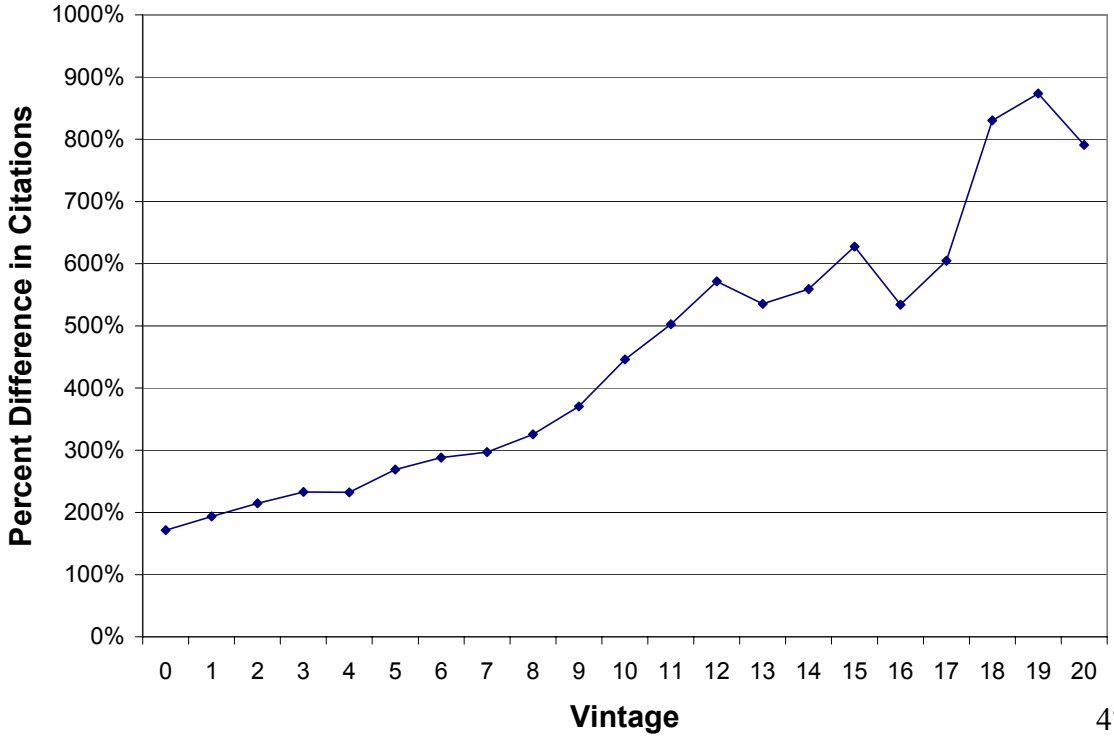
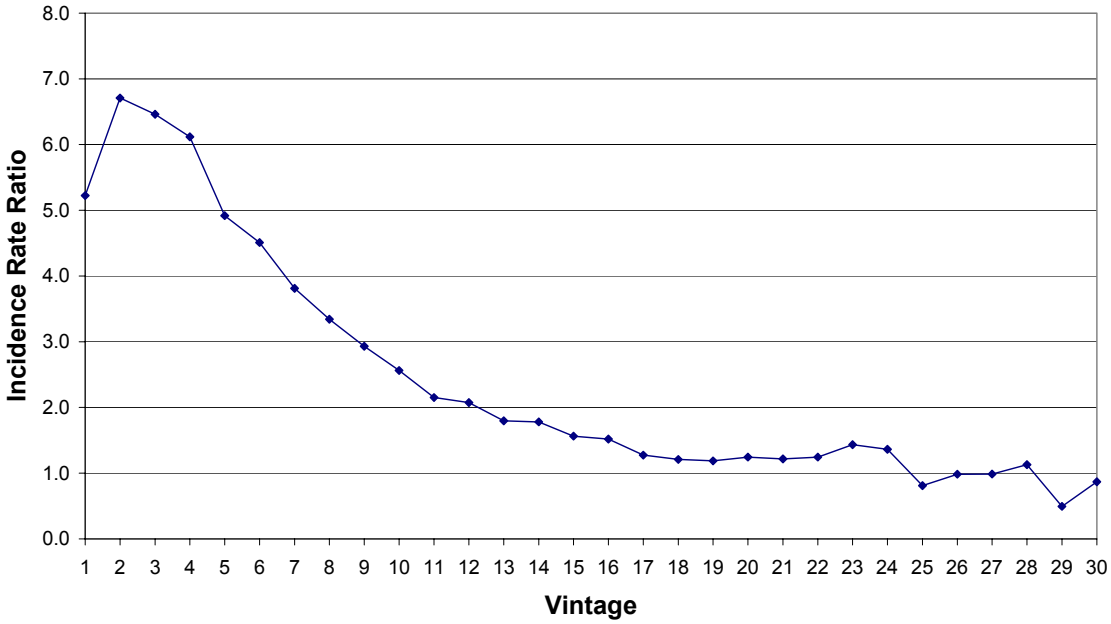


FIGURE D
CONTIDITIONAL VINTAGE EFFECTS



* Plot of Vintage Fixed Effects obtained in Negative Binomial estimation of CITED REFERENCES as a function of Article, Vintage, and Year Fixed Effects.

FIGURE E
DISTRIBUTION OF PAPER FIXED EFFECTS

FIGURE F
DISTRIBUTION OF DIFFERENCES IN NUMBER OF FORWARD CITATIONS, ATCC ARTICLES VS. CONTROL ARTICLES

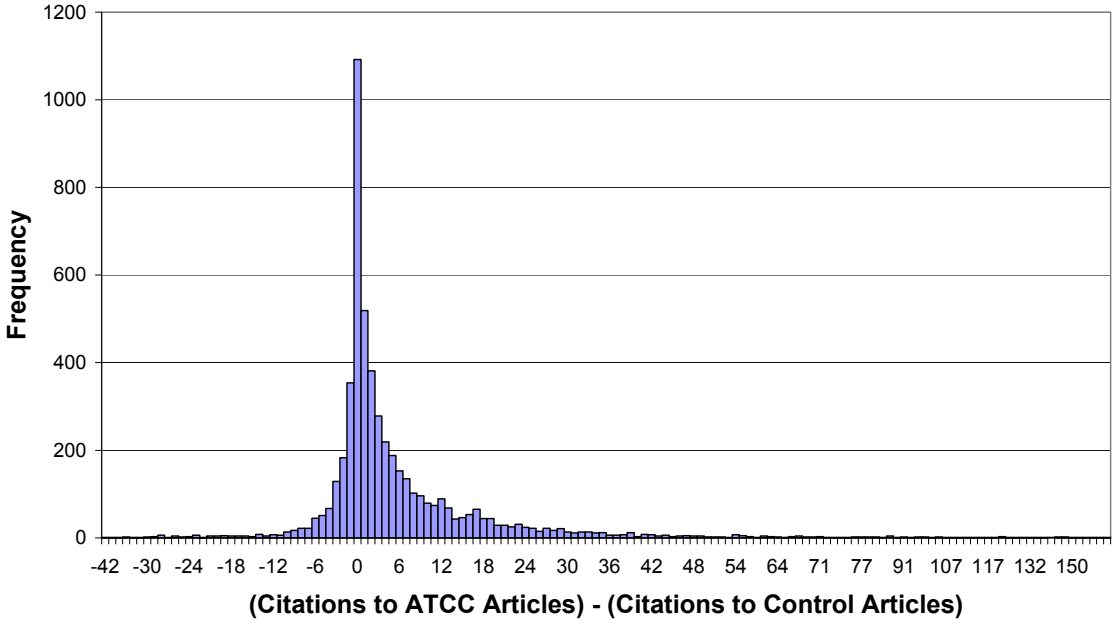


FIGURE G
IMPACT OF ATCC DEPOSIT ON FORWARD CITATIONS, BY COLLECTION

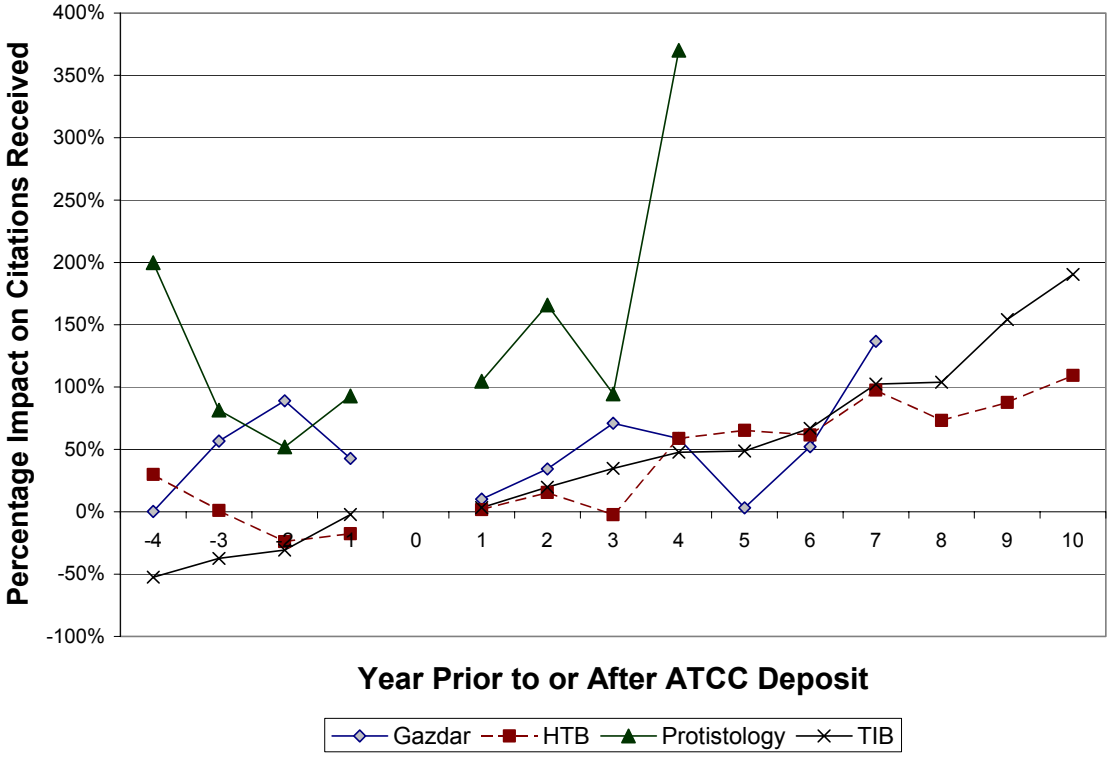
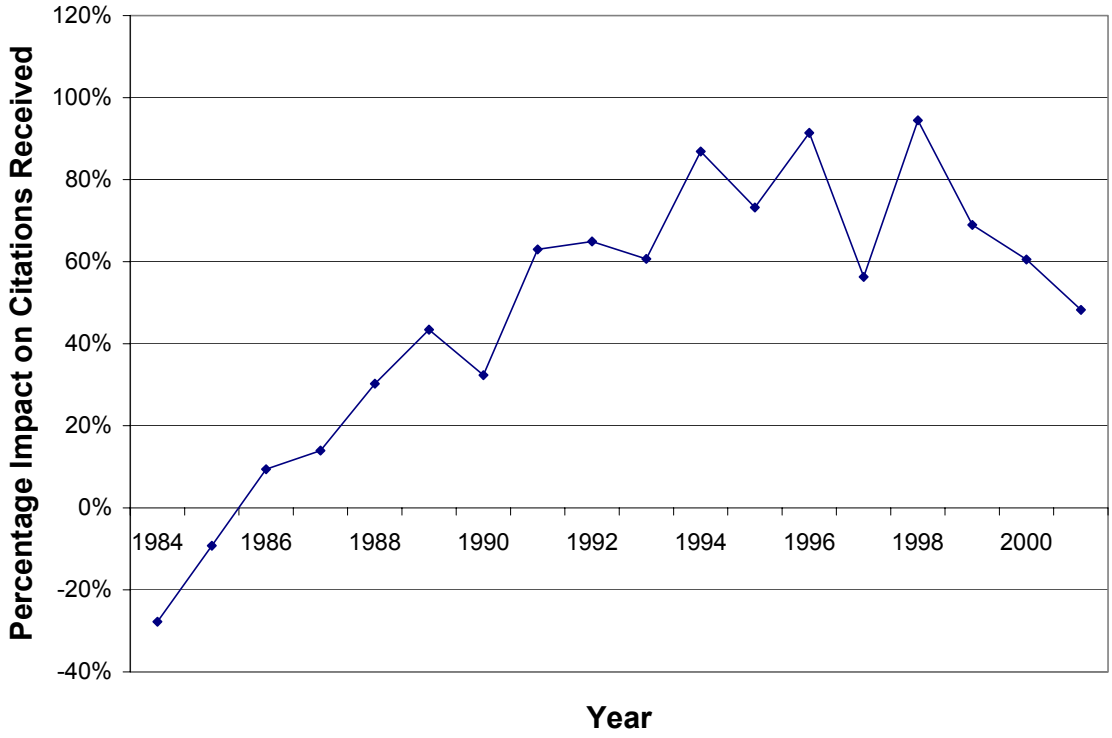


FIGURE H
IMPACT OF ATCC DEPOSIT ON FORWARD CITATIONS, MARGINAL EFFECTS BY YEAR



APPENDIX 1
PAPERS, BY COUNTRY

Australia	10
Belgium	7
Brazil	3
Canada	22
Denmark	3
France	16
Germany	24
Holland	4
Israel	2
Italy	8
Japan	26
Korea	1
Mexico	1
New Zealand	2
Poland	1
Russia	1
Scotland	3
South Africa	3
Spain	5
Sweden	6
Switzerland	7
Taiwan	1
Tanzania	1
United Kingdom	36
USA	396
Wales	1
Zimbabwe	1