CLIMBING ATOP THE SHOULDERS OF GIANTS: THE IMPACT OF INSTITUTIONS ON CUMULATIVE RESEARCH*

Jeffrey L. Furman
Boston University & NBER

Scott Stern
Northwestern University & NBER

November 1, 2007

* Acknowledgements: 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. We thank Andrew Bernard, Bronwyn Hall, Shane Greenstein, Ben Jones, Michael Kremer, Robert Litan, Megan MacGarvie, Fiona Murray, Paul Romer, Mathieu Trepanier and participants in numerous seminars for comments and suggestions. Jason Corradini, Mercedes Delgado, Lorraine DeLeon, Chijoke Emineke, Jeremy Gagne, Anna Harrington, Martha Kam, Julia Lo, Kinga Piekos, Courtney Mason, Tracy Myers, Michael Vitulli, and Naomi Wohl provided excellent research assistance. All errors are our own. Financial support for this research was provided by the Brookings Institution and by the Boston University School of Management Junior Faculty Research Fund. Author contact information: Scott Stern, Kellogg Graduate School of Management, 2001 Sheridan Road, Evanston, IL 60208 s-stern2@northwestern.edu, and Jeffrey L. Furman, Boston University School of Management, 595 Commonwealth Ave – #653a, Boston, MA 02215, furman@bu.edu.

CLIMBING ATOP THE SHOULDERS OF GIANTS: THE IMPACT OF INSTITUTIONS ON CUMULATIVE RESEARCH

ABSTRACT

While the cumulative nature of knowledge is recognized as central to economic growth, the microeconomic foundations of cumulativeness are less understood. This paper investigates the impact of a research-enhancing institution on cumulativeness, highlighting two effects. First, a selection effect may result in a high correlation between "high-quality" institutions and knowledge of high intrinsic quality and longevity. Second, an institution may have a marginal impact – an incremental influence on cumulativeness, conditional on the type and quality of knowledge considered. This paper distinguishes 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. BRCs may enhance the cumulativeness of knowledge by reducing the marginal cost to researchers of drawing on prior research efforts. We exploit three key aspects of the environment in which BRCs operate to evaluate how they affect the cumulativeness of knowledge: (a) the impact of scientific knowledge is reflected in future scientific citations, (b) deposit into BRCs often occurs with a substantial lag after initial research is completed and published, and (c) "lagged" deposits often result from shocks unrelated to the characteristics of the materials themselves. Employing a difference-in-differences estimator linking specific materials deposits to journal articles, we find evidence for both selection effects and the marginal impact of BRCs on the cumulativeness of knowledge associated with deposited materials. Moreover, the marginal impact increases with time and varies with the economic and institutional conditions in which deposit occurs.

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

I. INTRODUCTION

At least since the development of scientific societies and related research institutions in the 17th century, the centrality of cumulativeness in scientific and technical advance has been recognized.[1] However, from the perspective of economic theory, cumulativeness has only been incorporated recently, in models of endogenous economic growth [Romer, 1990; Grossman and Helpman, 1991; Jones, 1995] and step-by-step technical progress within industries [Scotchmer, 1991; Gallini and Scotchmer, 2003]. In order to serve as a foundation for long-term growth, scientific research and technological progress must exert a positive intertemporal spillover; as Jones [1995] emphasizes, to avoid diminishing returns to research investments, research itself must "stand on the shoulders" of prior knowledge.

Though extremely insightful in deriving the implications of cumulativeness for related economic variables (such as the equilibrium growth rate or the incentives for innovation), these models do not articulate the conditions that result in a cumulative research environment. For example, as Mokyr [2002] elegantly and persuasively argues, the mere production of knowledge does not guarantee that others will be able to exploit it. Effective diffusion of knowledge across researchers and over time requires that individuals are aware of extant knowledge and pay the costs of accessing that knowledge. Further, since any individual researcher only captures a small share of the benefit from the process of certifying knowledge and making it accessible, there may be a significant gap between the private and social returns associated with investments that contribute to the diffusion of scientific knowledge. Overall, then, the ability of a society to stand on the shoulders of giants depends not only on the amount of knowledge it generates, but on the quality of mechanisms for storing knowledge, the fidelity of knowledge, and the costs of accessing knowledge.

[1] Though certainly not the first example, Newton's recognition of cumulativeness is (famously) recognized in his classic 1676 letter to scientific rival 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." Economic historians and economists of technical change, most notably Nathan Rosenberg (1963), highlighted the centrality of cumulativeness in economic growth long before this idea was incorporated into formal models. As well, the role of institutions in promoting the explosion of scientific research in 17th century England is the cornerstone of Merton's seminal contributions to the sociology of science (Merton, 1957; 1973), which itself has served as a foundation for the "new" economics of science (Dasgupta and David, 1994).

Institutions and public policy are often suggested as central to the cumulative process. Social scientists face a considerable challenge, however, in assessing the extent to which any particular institution influences the way in which the "knowledge stock" is created, maintained, and extended. It is empirically difficult to isolate the impact of a particular piece of knowledge from the impact of the particular institution in which it is embedded, although the two are conceptually distinct. Two forces may be at work: First, a selection effect may result from a high correlation between "high-quality" institutions and knowledge of high intrinsic quality. However, for policy analysis, we are often more interested in the *marginal* impact of an institution – the incremental influence of an institution on cumulativeness, conditional on the nature and quality of knowledge embodied in it. Without an identification strategy capable of isolating selection and marginal effects, research examining the role of institutions on knowledge diffusion will confound these two effects.

The main contribution of this paper is to provide direct statistical evidence of the impact of a specific institution – biological resource centers (BRCs) – on the cumulativeness of knowledge, distinguishing the marginal impact of the institution from the effect of selection into that institution. Although this class of organization appears to be vitally important to research in the life sciences, their economic functions and contributions to knowledge spillovers have not received substantial attention from economists. 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 large and varied collections of biological materials, including cell lines, microorganisms, and DNA material, and distribute tools that allow researchers to access and exploit these materials. More than five hundred BRCs, cataloging more than 1.3 million deposits (principally bacteria, fungi, viruses, and cell lines) are registered with the World Data Center for Microorganisms [WFCC, 2006]. The American Type Culture Collection (ATCC), the largest BRC in the United States, distributes more than 250,000 materials annually [Stocker, 2006]; numerous smaller collections serve specific research communities, including the Plant Genetic Resources Conservation Unit (PGRCU), which annually distributes more than 35,000 seed, tissue culture, and clonal accessions [Pederson, 2005]. The apparent importance of such institutions to the life sciences derives from the importance of using certified research materials in life science research: The ability to build upon existing knowledge in the life sciences depends on access to the cells,

cultures, and specimens used in prior research (as well as certainty about the fidelity of those materials. BRCs constitute but one of a number of alternative institutional arrangements through which scientists can obtain materials for research purposes; however, the unique functions of these institutions in preserving, certifying, and circulating biomaterials materials have convinced numerous scientists of their fundamental importance to progress in the life sciences [Hunter-Cevera, 1996; OECD, 2001; Smith, 2003].

At a broad level, our analysis contributes to understanding the microfoundations of knowledge accumulation, which plays an important role in leading to ideas-driven growth. More specifically, our empirical approach extends recent studies using citation analysis to investigate the impact of institutions and technological communities on the cumulativeness of discovery and innovation [Jaffe, et al, 1993; Griliches, 1998]. We exploit three aspects of our empirical setting to develop and implement 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 that describes its initial characterization and application. Second, for specific types of BRC deposits, there is a significant lag between that initial article and its deposit into a BRC, and, in certain cases, the transfer of the materials was motivated by factors unrelated to the extent of their use. Specifically, we examine the deposit of materials into BRCs that are associated with "special collections" that are transferred *in toto* from smaller collections from which they had previously circulated into a major BRC, as a result of either the departure of a principal investigator or in response to institution-wide funding difficulties unrelated to the culture collection itself. Third, we take advantage of the fact that each of the special collections we analyze was a unified collection prior to deposit and was transferred as a group. Thus, the institutional shock is consistent for every material within a collection, and any systematic differences in the impact of that shock on future knowledge diffusion can be linked to the characteristics of the materials themselves.

In our empirical analysis, we focus on these "special collections" and evaluate whether articles associated with BRC deposited materials receive a boost in citations after deposit has occurred (after accounting for an article-specific fixed effect, and controlling for age and year fixed effects). In so doing, we are able to separately identify the role of selection (the likelihood that materials deposited into BRCs are associated with intrinsically important research) from the marginal impact of BRCs (the impact of BRCs in enhancing diffusion, controlling for the

intrinsic importance of that knowledge). Our approach builds on the considerable advances that have been made in recent years in evaluating differences-in-differences estimation [Bertrand, Duflo, Mullainathan, 2004]. Beyond our main specifications, we are able to provide a series of checks of our identification assumptions that reinforce our overall approach.

Our results provide strong empirical support for both the selection effect and the marginal impact of BRCs. Even in the period before their materials are accessible through a BRC, those research articles that are ultimately linked to BRCs experience nearly double the citation rate compared to a set of control articles drawn from the same journal and published in the same year. Even if the marginal impact of BRCs were to be zero, this result is important, as it suggest that the particular institution we study identifies and preserves materials that are, themselves, of greater than average importance in the life sciences. In fact, if the selection effect were not positive and significant, we may worry that a potential 'advertising' effect of BRCs may lead less important materials to be circulated. Conditional on materials becoming accessible through a BRC, their associated articles then experience a significant citation boost. The size of this boost ranges from just over 50% to more than 125% across the key specifications. In addition, this citation boost persists and, indeed, grows over time. This finding is consistent with the role of BRCs in helping to preserve the accessibility of knowledge for future research generations. While most articles are swiftly forgotten, the "rate of forgetting" associated with BRC-affiliated articles is substantially after accession. Taken together, these results suggest that, relative to alternative institutions, BRCs play an economically significant role in the intertermporal diffusion of knowledge in the life sciences.

To understand the mechanisms that drive this result, we further investigate whether the impact of BRC deposit depends on the economic or institutional conditions under which deposit occurs. We provide suggestive evidence that the benefits to BRC deposit arise *both* from a reduction in the "transactional" costs of accessing materials and an increase in the degree of certification associated with deposited materials. Specifically, we take advantage of the fact that different deposits in the same collection experience a similar "shock" in transaction costs, except for the differences in the prices charged by the BRC for different materials, and differences in the *ex ante* degree of perceived quality of those materials. Our results suggest that the impact of BRC deposit on future citations is lower for materials that are made available at higher prices,

and the citation boost is higher for articles associated with non-US authors (where the returns to certification may be higher).

The remainder of the paper proceeds as follows. Section II discusses the role of research-enhancing institutions in knowledge diffusion. Section III describes BRCs, focusing on the mechanisms by which these institutions lower the cost and increase the "tightness" of knowledge over time. Section IV outlines a differences-in-differences framework for identifying the impact of BRCs on knowledge diffusion. Sections V and VI review the data and present the empirical results, respectively. A final section concludes.

II. THE IMPACT OF RESEARCH-ENHANCING INSTITUTIONS ON THE DIFFUSION OF KNOWLEDGE

The dynamic accumulation of knowledge has become a central issue to many different areas of economic research. Institutions and public policy are often suggested as key determinants of the ability of an economy to sustain cumulative knowledge production [Rosenberg, 1963, 1979; Heller and Eisenberg, 1998; David, 2001; Mokyr, 2002]. The diffusion of knowledge over time depends on institutions that facilitate low-cost knowledge transfer among researchers and over research generations. Institutions may lower the costs of access to useful knowledge by enhancing "the technology of access, the trustworthiness of the sources, and the total size of the [stock of knowledge about the natural phenomena and regularities]" [Mokyr, 2002, p. 8].[2] We refer to economic institutions that promote the cumulativeness of the research process (through one or more of these mechanisms) as *research-enhancing institutions*.

Over the past two decades, a great deal of qualitative and quantitative economic research has investigated specific research-enhancing institutions, often with the objective of documenting the presence of knowledge spillovers.[3] While the attempt to identify and measure knowledge production and diffusion is inherently difficult [Griliches, 1990], a sophisticated empirical literature has emerged recently that attempts to identify the impact of particular

[2] Put another way, "Progress in exploiting the existing stock of knowledge depends first and foremost on the efficiency and cost of access to knowledge" (Mokyr, 2002, p. 7). While we focus on the role of formal knowledge-sharing institutions, substantial (and ever-increasing) human capital investments in specialized scientific and engineering knowledge are perhaps the single most important barrier to discovery at the frontier (B. Jones, 2003).

[3] While systematic empirical evidence is more recent, the linkage between institutions and cumulativeness has been emphasized at least since Vannevar Bush's 1945 policy manifesto, *Science: The Endless Frontier*. Nelson (1959) and Arrow (1963) built on Bush's compelling articulation of the role of basic research in economy-wide prosperity to identify the public goods nature of basic research and the case for public investment. More recently, the national innovation systems literature (as pioneered by, among others, Nelson, 1993) emphasizes the role of research-enhancing institutions in mediating geographically-localized knowledge spillovers.

institutions on the extent of knowledge spillovers. This research often employs citations to academic papers or granted patents to estimate the influence of prior knowledge on current advances.[4] Perhaps no research-enhancing institution has been more intensively studied than the university. For example, Jaffe et al [1993; 1998] examine whether university patents receive citations at a significantly higher rate and with significantly greater geographical scope than a group of "control" patents drawn from similar geographic and technological areas. More recently, Branstetter [2003] reviews patterns of patent citations to academic research papers, and finds that spillovers from academic science to commercialized inventions occurs in a limited set of technological fields and geographic areas. As well, prior studies have investigated the role of specific policies, such as the Bayh-Dole Act or the strengthening of patent rights in Japan [Mowery et al., 2001; Mowery and Ziedonis, 2002; Sakakibara and Branstetter, 2001]. The "search for spillovers" has extended its reach beyond university and IP law, now including studies of R&D consortia [Irwin and Klenouw, 1996; Branstetter and Sakakibara, 2002], the national laboratories [Jaffe and Lerner, 2001), venture capital (Kortum and Lerner, 2000], and patent pools [Lerner and Tirole, 2003], among others.[5]

While this prior literature has established a close empirical association between research-enhancing institutions and the impact of scientific and technical knowledge (as reflected in higher rates of citations to papers and patents, respectively), prior research has not been able to disentangle whether these institutions facilitate cumulativeness *per se*, or whether they are simply linked to knowledge which has a higher intrinsic impact. In the terminology of the program evaluation literature, prior research conflates the *selection* effect (high quality institutions are simply associated with high quality knowledge) with the *marginal* impact of those institutions on knowledge diffusion. For example, university patents may be more highly cited (relative to a control group of patents generated by industrial-based laboratories) because the research reflected in the patent is more fundamental or because the norms of disclosure and

[4] While citations are certainly not the only means by which we can measure the cumulative impact of a given piece of scientific research, citations are a useful though noisy indicator of the exploitation of knowledge by subsequent researchers (and are likely more informative in the life sciences than in social science disciplines such as economics or sociology). Most research papers in the life sciences are short and focused, with few extraneous references to literatures beyond those directly impacting the specific results described. As a result, the principal rationale for the inclusion of a citation for a paper associated with a BRC deposit is when the material is explicitly used in a follow-on experiment, or the experiment is closely connected to the research findings and knowledge linked to that specific material.

[5] It is useful to note that a sociological literature has also developed, focusing on whether the ability of a researcher to draw upon others' knowledge is linked to their participation and position within specific social networks in which that knowledge is embedded (Powell, 1998; Rosenkopf and Tushman, 1998) and on the norms of use associated with that knowledge (Sorenson and Fleming, 2004).

openness associated with a university contribute to more effective diffusion of that knowledge. In other words, the long-term impact of knowledge depends not only on its importance but on its linkage to institutions that facilitate low-cost knowledge diffusion. The remainder of this paper is devoted to disentangling these two effects in the context of a specific research-enhancing institution, biological resource centers.

III. BIOLOGICAL RESOURCE CENTERS AND CUMULATIVE RESEARCH IN THE LIFE SCIENCES[6]

III.A. The Problem of Authenticated Biomaterials & the Advance of Life Sciences

Compared with many other scientific and technological areas, research in the life sciences has developed dramatically in the last several decades. While scientific opportunity has likely played an important role in these advances, the life sciences have also benefited from their ability to address a number of the constraints that may otherwise limit step-by-step progress. One of the central constraints on the capacity of life science research to advance knowledge is its ability to maintain the integrity of shared biomaterials. Because lines of inquiry in life sciences require building on the advances of others, researchers often must rely on research materials initially generated in other laboratories. Equally, follow-on research even within the same laboratory or set of labs requires exact replication of research materials used in prior experiments. Without an effective system to ensure that the materials used in particular experiments are exactly the materials that researchers believe they are using, advances in microbiological research would be limited.

Though seemingly simple, this straightforward problem has bedeviled the life sciences research community since the advent of transportable biomaterials in the 1950s. Before researchers grasped the importance of biomaterials fidelity (and before verification techniques were developed and widely understood), it was the norm for researchers to exchange biomaterials through peer-based networks. As a consequence, a surprisingly large share of the materials exchanged between labs became contaminated with other cell lines or otherwise misidentified. Precisely such a problem cast doubt over thousands of individual research findings, including the research of Nobel Prize-winning scientists and other researchers around the world in series of revelations in the 1970s that became known as the "HeLa scandals" [Gold,

[6] Stern (2004) provides a more thorough description of the functions and history of BRCs. See, also, Cypess (2003) and OECD (2001) for an introduction to BRC functions and policy issues.

1986]. Although the most well-known cases of misidentification were uncovered during the 1970s, some researchers argue that a substantial fraction (more than 20%) of current cell lines are misidentified [MacLeod, et al, 1999] and that thousands of published articles each year employ such misidentified cell lines [Masters, 2002].

In addition to concerns about the prospect that mischaracterized materials may yield errors in research, uncertainty about the fidelity of biomaterials alone can result in considerable research delays, as scientists must undertake substantial efforts to verify each of the materials they employ. Thus, the rate of advance in the life sciences depends critically on researchers' confidence in the authenticity and purity of their materials.

It is important to emphasize that the problem of maintaining the fidelity of research materials is not principally a technological or scientific problem but is driven by the economics of research incentives. Whereas individual scientists have few incentives to engage in replication and validation activities, researchers who have published specific findings may find it worthwhile to limit scrutiny of their results. Though a robust system for validating experimental research is in the interest of all scientists, individual researchers have few incentives to contribute to this outcome. Because the integrity of the scientific process is a public good, an institutional response is crucial for addressing this problem.

III.B. Institutional Arrangements for Sharing Biomaterials

Several alternative institutional arrangements exist for collecting, certifying, and circulating biological materials, including peer-to-peer networks, for-profit culture collections, private culture collections, and biological resource centers.

Peer-to-peer networks consist of informal exchanges among researchers and are dependent on research laboratories maintaining culture collections and fulfilling requests by other researchers for distribution. In some cases, researchers may maintain small personal collections of materials used in current work; in other cases, researchers maintain modest-sized collections of materials that they store and exchange regularly with colleagues within and across their institution. In a "pure" peer-to-peer network, it is impossible to require researchers to exchange materials and initial discoverers may be reluctant to offer access to researchers whose experiments could undermine the value of the initial work. At the very least, peer-to-peer transactions require that researchers who want to use particular research materials contact, and,

occasionally, negotiate with the initial developer of the material. In some cases, negotiations over access to materials, such as lab data, or cell lines may be conditional on the recipient extending an offer of co-authorship or some other incentive. In other cases, the transfer of materials may be dependent upon the recipient agreeing to the terms of a Materials Transfer Agreement. The problem of verification may also be substantial in peer-to-peer networks, as the scientists making and agreeing to exchanges often rely on layers of assistants to prepare, send, receive, and employ research materials.[7]

Some laboratories maintain culture collections than are larger than those usually maintained for personal use by individual researchers. These collections circulate materials on a fairly regular basis, generally without formal restrictions such as those imposed by Materials Transfer Agreements. Although they circulate their materials fairly regularly, these collections have limited resources to dedicate to preservation and certification functions. Although they face fewer restrictions on materials circulation, these larger academic collections are subject to some of the same difficulties in verification as are individual collections in the peer-to-peer network.

For-profit biomaterials distribution firms and private collections, such as those maintained by major pharmaceutical companies, constitute alternative institutions for circulating biomaterials. Not surprisingly, for-profit firms tend to "cherry-pick," focusing on a narrow range of materials that offer high margins and low storage costs. Reputation-based mechanisms and the relationship between quality and sales may lead their materials to be relatively trustworthy. Proprietary collections, such as those maintained by companies such Merck and Lilly, are designed to benefit individual firms. However, since circulating catalogs of their materials would communicate competitive intelligence to rivals, these organizations rarely circulate their materials externally and require extensive Materials Transfer Agreements when they do.

III.C. What are Biological Resource Centers?

Biological Resource Centers constitute an additional alternative to peer-to-peer networks private collections, and for-profit biomaterials distribution firms. BRCs are institutions that collect, certify, and distribute biological organisms for use in life science research and in the

[7] It is possible that informal "brokers" will emerge in a peer-to-peer network, facilitating transactions in the "market for ideas" (Lamoreaux and Sokoloff, 1999; Gans and Stern, 2002). However, brokers are limited by their incentives and the extent of their personal networks. As well, the connections enabled by a broker are subject to termination if the broker discontinues her role (because of retirement or changing interests). The fact that it is difficult to identify precisely who is responsible when shared materials become contaminated also complicates the prospect that reputation-based mechanisms will ensure that researchers share their "best" materials even when asked.

development of commercial products in the pharmaceutical, agricultural and biotechnology industries – in a sense, they serve as "living libraries" preserving and circulating research materials. 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. Indeed, life scientists are in broad agreement that such institutions play a vital role in long-term scientific progress [Hunter-Cevera, 1996; OECD, 2001; Smith, 2003].

The definition and scope of BRCs have evolved over time. Initially referred to as culture collections or cell banks, experts described their functions as "acquiring, preserving, authenticating, and distributing microorganisms or in vitro cultured cells to qualified scientists" [Stevenson and Hatt, 1992]. BRCs are therefore related to but distinct from mammalian research supply institutions, such as Bar Harbor's Jackson Laboratory, which provides JAX knock-out mice to researchers throughout the world [Murray, 2005]. Over the past few decades, the scope of traditional culture collections have expanded, leading most centers to reclassify themselves as biological resource centers. [8]

Since the 1980s, select BRCs have been critical to the extension of intellectual property rights, by serving as international patent depositories for all patented living organisms. Indeed, according to the terms of the Budapest Treaty for the Deposit of Cultures that are the Subject of Patent Procedures, which was introduced in 1977 and implemented beginning in 1981, all patented biological materials must be deposited in a certified BRC in order for that material to receive patent protection.[9]

BRCs can differ in their functioning and governance. As of 2006, the World Federation of Culture Collections lists more than 520 distinct BRCs around the world. Collections vary dramatically in the size of their holdings, ranging from less than 100 to more than 100,000 cultures. Overall, more than 1.3 million cultures are maintained by BRCs throughout the world, and more than a million of cultures are distributed each year by BRCs to scientists throughout

[8] In a recent report, the OECD describes BRCs' as, "service providers and repositories of living cells, genomes of organisms, and information relating to heredity and the function of biological systems. BRCs contain collections of culturable organisms (e.g. micro-organisms, plant, animal and human cells), replicable parts of these (e.g. genomes, plasmids, viruses, cDNAs), viable but not yet culturable organisms, cells and tissues, as well as databases containing molecular, physiological and structural information relevant to these collections and related bioinformatics" (OECD, 2001).

[9] See Stern (2004, p. 52-53) for an overview of the Budapest Treaty.

the world. Large national collections, such as the ATCC in the United States or DSMZ in Germany, manage a broad range of materials, while specialized collections, such as Yale University's E. Coli collection, provide a niche resources for particular scientific areas. Although all collect, identify and distribute materials, BRCs vary in how exhaustively they authenticate and certify the materials that constitute their collections. For example, while leading BRCs in the United States such as ATCC and the Coriell Institute are both recognized for their international leadership in cell culture technique, newer and smaller collections have tended to adopt techniques and protocols developed elsewhere. Finally, while the plurality of BRCs around the world rely principally on government financing, some large national collections, especially those in the United States, rely on a mixture of public and private support. For example, the ATCC is organized as a private not-for-profit institution, which derives the majority its income from activities such as its patent depository distribution services, although it also receives governmental support. (Appendix Table I lists examples of various types of Biological Resource Centers.)

Despite some differences in form, BRCs share a common purpose: they serve as repositories that make materials and research results developed by one generation of researchers available for future research endeavors. In so doing, prototypical BRCs exhibit characteristics of the type of institutions that Mokyr [2002] describes as potential contributors to cumulative knowledge growth. In contrast to alternative institutional arrangements, archetypal BRCs engage in the preservation and certification of biomaterials, offer independent access to those materials, and take advantage of economics of scale and scope in the process. While some alternative institutional arrangements support some of these activities – for-profit collections also invest substantially in in-house certification, for example – only BRCs are dedicated to the full range of these activities. As a result, well-functioning BRCs may be able to lower the cost of follow-on research substantially in comparison to alternative institutional arrangements.

III.D. Features of Biological Resource Centers

III.D.1 Preservation of Biological Materials

The *preservation* of biological materials is a primary function of BRCs. BRCs collect, characterize, and maintain an exceptionally broad collection of biological materials, including materials whose value is not initially understood. For example, the largest collection in the

United States, maintained at the American Type Culture Collection (ATCC), includes more than 92,000 strains of micro-organisms and cell isolates, and more than 5,000,000 DNA sequences. In Europe, the large German collection, the DSMZ, maintains more than 16,000 cell cultures, representing more than 5,000 distinct microorganism species. BRCs retain these collections over extremely long periods of time, even when specific applications are not immediately apparent.

The potential windfall associated with long-term preservation can be seen through illustration. Consider the case of *Thermus aquaticus*. In the late 1960s, Thomas Brock discovered a new type of micro-organism – an *extremophile* – which lived in the rather severe climate of the hot springs of Yellowstone National Park. Though no practical benefit was foreseen at the time, *extremophiles* exhibited a number of distinctive properties, including the ability to sustain enzymatic reactions during rapid heating and cooling (a property which was critical for sustaining life in a geyser). At the time of its discovery, the ATCC then established and maintained an *extremophile* collection for use by both academic and industrial researchers. More than 15 years later, Kary Mullis, a researcher at Cetus Corporation, experienced a fundamental insight when he conceived of polymerase chain reaction (known as PCR), the basic technique that allows for rapid replication of DNA (at the time, this was arguably the single largest bottleneck in biotechnology research). Mullis' insight required a material that could withstand extreme temperature variation during the replication process. While the effort to develop such a material synthetically might have taken years, Cetus researchers were able to identify and draw upon the *extremophile* collection at the ATCC to rapidly implement a practical approach to PCR, revolutionizing the modern life sciences. Not only did PCR dramatically improve research productivity in the life sciences (e.g., resulting in the core techniques behind the Human Genome Project), but Mullis shared the Nobel Prize in 1993, and *Thermus aquaticus* itself was named "Molecule of the Year" by *Science* in 1989.

BRCs offer particular advantages in preservation relative to alternative institutional arrangements. For example, because the IP rights held by for-profit laboratories exists for only a modest time (often less than the time between initial characterization and greatest potential use), the for-profit community has few incentives to maintain the widest range of materials indefinitely. Indeed, for-profit distributors of biological materials tend to "cherry-pick" a narrow range of materials offering high margins and low storage costs. As well, private life science firms, such as pharmaceutical companies, often maintain their own in-house facilities. These

facilities are likely to preserve a narrow range of materials relevant to their own research efforts, and only for periods corresponding to their expected in-house use. Compared to traditional laboratories that maintain most materials for less than a decade, BRCs have established procedures and technologies to allow materials to be preserved for many decades (and even perhaps for centuries). While private not-for-profit collections exist in large numbers within the peer-to-peer scientific network (most of which are highly duplicative of each other), most collections are narrow, and depend on the idiosyncratic interest (and unpaid effort) of individual researchers, raising the possibility that materials will be lost due to retirement or inattention by culture curators.[10] Whereas BRCs explicitly focus on preservation, there are few incentives within for-profit entities or the peer-to-peer network to maintain a full range of materials for an indefinite period of time. By serving to facilitate large scale retention and maintenance of biological materials, BRCs both aid knowledge diffusion in the short term and limit costly duplication effort over time.

III.D.2 Certification in Biological Resource Centers

BRCs also *certify* research materials. While BRCs do not fully replicate published experiments, materials incorporated into BRC collections undergo a series of reviews and tests to establish their identity and biological viability. BRCs therefore provide the means for scientific replication. Some BRCs, such as the ATCC, offer a classification system that allows researchers to evaluate the degree of confidence associated with specific deposits. The ATCC and DSMZ, for example, regularly issue notices identifying materials errors and misclassifications.

As illustrated in the case of massive HeLa contaminations, the consequences of misidentification are far-reaching.[11] Not only does misidentification cast a cloud over the findings of current researchers (with career implications for those whose results are under suspicion), but confusion and uncertainty places a longer-term cost on progress. In addition to

[10] For example, in January, 2002, three private university collections were identified as "orphans" available for new storage site; two of these three were classified as "defunct" by July, 2002 (methanogens.pdx.edu/usfcc).

[11] Unfortunately, the HeLa case is not simply an isolated, historical mistake. Misidentification of biological materials plagues published (and patented) research findings to this day (Masters et al, 2001). The story of the KB cell line is a recent example. Originally derived from oral cancer cells, KB became contaminated with HeLa cells. In other words, it is well documented that researchers using the KB line are actually performing experiments with the HeLa line. Despite this, more than 300 published articles based on the KB line were published between 1998 and 2000, many of which claimed to provide new findings specific to oral cancer (Masters, 2002). Moreover, many of these articles have themselves been extensively cited by subsequent researchers. The persistence of misidentification is a consequence of the incentive system in scientific and commercial research: high-powered incentives to claim priority over a novel discovery, and few if any incentives for individuals to validate research claims made by others.

minimizing the likelihood of contamination and spurious findings, effective certification enables researchers to avoid needless and costly duplication, and thus increase research productivity over time. In the absence of effective certification procedures, researchers must painstakingly re-establish the validity of specific findings in order to design and implement new research: they must literally re-invent the wheel. As highlighted by Mokyr [2002], the "tightness" of knowledge is crucial for the effective use of knowledge; certification by an "invisible" institution such as BRC enhances the tightness of knowledge and so allows researchers to increase their productivity by avoiding costly verification procedures.

Relative to peer-to-peer materials distribution systems, BRCs possess significant advantages in certifying biomaterials. First, the repeated nature of BRC-scientist transactions enhances the incentives that BRCs face to ensure the fidelity of all materials. As well, larger institutions are more capable of bearing the fixed costs associated with materials certification, which include the costs of substantial laboratory equipment, information technology, and skills development.

III.D.3 Independent and Open Access to Biological Materials

Third, BRCs advertise the availability of materials in their collections and ensure that these are equally accessible to all members of the scientific and technological community, thus encouraging *independent and open access* to the results of prior scientific research. In non-BRC networks, access to source materials is dependent on the "goodwill" 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 follow-on research 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 databases and materials; recent controversies, such as that over the ownership and use of the Harvard Oncomouse, licensed to DuPont, reflects an ongoing and recently pitched battle over access to biological materials and data [Murray, 2005].

Building on earlier research in the peer-to-peer system may involve protracted negotiations with the initial scientists (e.g., about coauthorships or intellectual property claims). Independent access to research materials is required for replication and so is at the heart of the

scientific method in biological and medical research. However, the incentive for individual scientists to grant access is limited within the modern university and private life sciences research environment. Even after results are published (and perhaps because they are published), researchers may hold up efforts by others to gain access to materials, both in order to further their research lead and to avoid detailed investigation of their research conclusions.

Consider the costly controversy over the discovery of the AIDS retrovirus. The race to discover the cause of AIDS involved an intense and competitive battle between French researchers at the Institut Pasteur and Dr. Robert Gallo's lab at the National Cancer Institute. Though the French team first isolated the correct virus, laboratory-to-laboratory material exchanges resulted in nearly a decade of confusion about the precise nature of the virus and the allocation of credit for its initial discovery. These incidents damaged Gallo's reputation and delayed critical AIDS discoveries. At least in part, delays in discovery resulted from insufficient incentives for individual laboratories to provide low-cost, independent access to their own research materials.

In contrast, BRCs sever the direct tie between the researcher associated with an initial discovery and those want to build upon the research. Materials available in BRC collections are listed either on public websites or in catalogs. Relative to a private collection or the peer-to-peer network, BRCs lower the costs of accessing research materials. The importance of this is non-trivial: a great deal of knowledge consists of "knowing that something is known and knowing how to find it" [Mokyr, 2002, p. 9].

III.D.4 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, 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. These scale and scope economies are reflected in the use of non-profit BRCs by

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. By achieving economies of scale and scope, BRCs can lower the transaction cost of access to the existing stock of knowledge.

IV. THE IMPACT OF INSTITUTIONS ON KNOWLEDGE DIFFUSION: AN EMPIRICAL FRAMEWORK

By ensuring the fidelity of and lowering the costs of access to knowledge, institutions such as BRCs may influence the equilibrium rate and impact of a given discovery on subsequent research. Three central predictions stand out. First, conditional on its intrinsic scientific importance and quality, a discovery linked to a research-enhancing institution will have a higher diffusion rate, relative to the case where such knowledge was produced and diffused independently of such an institution. Second, the marginal impact of association with a research-enhancing institution will increase over time. Research-enhancing institutions preserve access to discoveries and knowledge for a much longer period of time than is feasible under alternative institutional arrangements. Third, the selection effect suggests that knowledge associated with BRC materials may tend to have a higher (or lower) intrinsic scientific value than knowledge associated with materials diffused through the peer-to-peer network.

The presence of a selection effect results in a fundamental inference problem. Specifically, for a given piece of knowledge produced or diffused within a given institutional environment, one cannot directly observe the counterfactual impact that knowledge would have had if the knowledge had been produced and diffused in an alternative institutional setting. For example, if researchers and BRCs endogenously acquire biological materials tending to have high fundamental scientific interest, a simple comparison of the impact of knowledge linked to a BRC versus knowledge with no BRC linkage will be biased. From an experimental perspective, the econometrician would ideally observe a given piece of knowledge in distinct institutional environments and compare the impact of that knowledge across regimes.

While one cannot replicate this ideal experimental design, this paper develops and implements an econometric strategy that takes advantages of the institutional environment to

estimate the role of selection and marginal effects in the diffusion of scientific knowledge. Our approach exploits two key elements of the system by which scientific research is diffused. First, individual materials made available through BRCs are linked to specific scientific publications. We can therefore assess the impact of BRCs by examining the pattern of citations to articles associated with BRC deposits. Though imperfect, citations by future scientific research articles provide a useful (though noisy) index of the "impact" of a discovery on subsequent research.

Second, many BRC material deposits occur long after the publication date of the associated scientific research article; moreover, in a number of instances discussed in the next section, the act of deposit and its precise timing are arguably econometrically exogenous (and we can apply differences-in-differences techniques to test whether, indeed, these deposits are exogenous). We therefore exploit the timing of transfer for some collections that had been maintained in academic laboratories that get shifted into a public BRC (e.g., when the principal investigator retires or switches university affiliation). In other words, while initial publication often occurs within six months (or fewer) after initial journal submission, there are often substantial delays between initial publication and BRC deposit. For scientific research articles linked to BRC deposits that occur with a lag, we thus observe both a pre-deposit and post-deposit period. This allows us to estimate the impact of deposit on knowledge diffusion, measured as the change in the rate of citation to the initial article by follow-on scientific research articles.

By linking BRC deposits to potentially citable scientific research articles, we implement a differences-in-differences estimator of the marginal impact of BRC deposit.[12] Specifically, we construct a dataset composed of scientific publications linked to (delayed) BRC deposits and two separate groups of control articles, each of which is comparable to our treatment articles in terms of *ex ante* expectations of scientific impact. (We describe these in greater detail in Section IV.) Because we observe citations to a scientific publication both before and after BRC deposit (and because we observe control publications never linked to BRC deposits) we are able to identify how the pattern of citations to a scientific publication changes as the result of BRC deposit. This test goes beyond the potentially biased test of whether BRC-linked articles are more or less highly cited than those that are not associated with BRC deposits.

[12] We discuss our identification argument in more detail in Section V. It is useful to note that we also check whether the timing of deposit is exogenous by testing for the presence of a pre-BRC deposit trend that "predicts" the act of BRC deposit. As discussed in Section VI (Figure E), our results are robust to the inclusion of such a trend, and we do not find statistically significant evidence of such a trend.

More precisely, if the availability of research materials through a BRC lowers the cost and raises the expected value of building on a specific research contribution, then the citation rate to BRC-linked scientific publications should increase after deposit has occurred. Of course, measuring the impact of scientific research using citations implies that we must account for its form as count data that are skewed to the right (and likely over-dispersed relative to Poisson). Therefore, except where noted, we employ a negative binomial model of the citations produced per year for each scientific article in our dataset. As well, the rate of citation to a given piece of research will vary with the calendar year, with the time elapsed since initial publication and across different article "families" (where a family is composed of a BRC-linked article and the two control articles). Except where noted, the empirical specifications account for these effects through the use of age, year and family (or article) fixed effects.[13,14]

To disentangle the relative role played by selection versus the marginal impact of BRC deposit, our analysis first considers an estimator that identifies the average difference across the treatment and control groups, and estimates the change in citations resulting from BRC deposit itself. Specifically, this baseline estimator is simply:

$$(1)\quad CITES_{i,j,pubyear(j),t} = f(\varepsilon_{i,j,t}; \alpha_j + \beta_t + \delta_{t-pubyear} + \phi BRC_i + \psi POST-DEPOSIT_{i,t})$$

where α_j is a fixed effect for each article family, β_t is a year effect, $\delta_{t\text{ - pubyear}}$ captures the age of the article, BRC is a dummy variable equal to one for those article linked at some point to a BRC, and POST-DEPOSIT is a dummy variable equal to one only for years after the material linked to the article is accessioned and available from a BRC.[15] While this specification provides an estimate of relative importance of the selection effect and the marginal impact of BRC deposit, the potential for substantial heterogeneity among articles (even within article families) may lead to an upward estimate of the impact of BRC deposit on subsequent citation. We

[13] Several subtle issues, including the incidental parameters problem, arise in incorporating multiple fixed effect vectors into a negative binomial specification. We have experimented with a range of alternative procedures and approaches, including the conditional negative binomial estimator suggested by Hausman, Hall, and Griliches (1984) and the fixed effects estimator suggested by Allison and Waterman (2002). Our core results are based on the traditional conditional fixed effects estimator with bootstrapped standard errors; however, our qualitative findings are consistent across these different procedures.

[14] When using a conditional fixed effects estimator, one citation year and one age fixed effect are not separately identified (Hall et al, 2005). Since the main effect that we are interested in is separable from these effects, the precise specification we employ to overcome this identification issue does not at all affect our estimate of the impact of BRC deposit on citations. In our estimation, we identify differences relative to age = 0, and relative to publication in years after 1975 (though, due to data limitations, we actually impose a single regressor on the years 1975-1979).

[15] Our empirical specifications also incorporate a "window" including the year prior to and year after the accession of a material into the BRC to account for "announcement effects" and for potential lags in availability of materials.

therefore examine (and base our core findings on) a series of estimates including article-specific fixed effects (γ_i), as in the following specification:

$$(2)\quad CITES_{i,j,pubyear(j),t} = f(\varepsilon_{i,j,t}; \gamma_i + \beta_t + \delta_{t-pubyear} + \psi POST-DEPOSIT_{i,t})$$

Overall, we test for the impact of research-enhancing institutions by calculating how the citation rate for a scientific publication *changes* after BRC deposit, accounting for fixed differences in the citation rate across articles and relative to the non-parametric trend in citation rates for articles with similar characteristics.

To test the preservation hypothesis, we can estimate whether the impact of BRC deposit changes with the time elapsed since BRC deposit itself. As well, we can check for the presence of a pre-deposit time trend (which might argue against the exogeneity of the deposit event itself). We simply modify (2) to allow for pre-deposit and post-deposit dynamics:

$$(3)\quad \begin{aligned} CITES_{i,j,pubyear(j),t} = f(\varepsilon_{i,j,t}; \gamma_i + \beta_t + \delta_{t-pubyear} + \\ \sum_{k=1....10} \psi_{PRE\_k} PRE-DEPOSIT(k)_{i,t} + \sum_{l=1....10} \psi_{POST\_l} POST-DEPOSIT(l)_{i,t}) \end{aligned}$$

where PRE-DEPOSIT(k) and POST-DEPOSIT(l) are dummy variables equal to one in the year when a BRC-linked article is a given number of years prior to or after the deposit event. Concerns about endogeneity can be tested by examining whether the coefficients on ψ_{PRE_k} increase in the few years prior to the initial announcement of BRC involvement, and the preservation hypothesis can be tested by whether ψ_{POST_l} is increasing over time.

Finally, we can use interaction effects to investigate whether the impact of BRC deposit depends on the economic or institutional conditions under which deposit occurs. For example, our identification strategy focuses on the transfer of three distinct biological materials collections that were shifted from the peer-to-peer network to the ATCC, the leading BRC in the United States. By interacting POST-DEPOSIT with a dummy variable for each of these collections, we can separately identify whether the impact of BRC deposit varies across different deposit "events." This same strategy can also be used to evaluate the mechanism underlying the impact of BRC deposit. Conditional on being associated with one of the deposit "events," the impact of the transfer from the private collection to the BRC differs only in terms of (a) the price charged by the BRC for access to the material and (b) the *ex ante* degree of perceived quality of those materials. By interacting a price variable and measures of the "reputation" of the article or

author prior to deposit with the POST-DEPOSIT measure, we are able to evaluate how the returns to BRC deposit vary with changes in the transactional costs of access and the returns to research validation.

V. DATA

V.A. Data Construction and Sources

To conduct this empirical analysis, we focus on a single institution, the American Type Culture Collection (ATCC). Located in Manassas, Virginia, and founded in 1925, the ATCC maintains the largest culture collection in the world. Although the ATCC is of unusually large size, its preservation, certification, and distribution functions are similar to those of other national and specialized public culture collections, such as the DSMZ in Germany and the Japan Collection of Microorganisms, and the Coriell Medical Research Institute and the Agricultural Research Service Culture Collection (NRRL) in the United States.

In addition to being representative of major culture collections, we are able to take advantage of the characteristics of ATCC in order to address four key empirical challenges associated with implementing the differences-in-differences strategy we describe above: (a) linking BRC deposits to research publications, (b) identifying a sample of publications that can be used to disentangle the impact of selection versus the marginal impact of BRCs, (c) constructing a sample of control articles, and (d) accounting for ambiguity in the date at which BRC deposits are available for access by other researchers.

We address the first challenge by taking advantage of the reference information maintained by the ATCC on all materials deposited in its collections. For each material, ATCC documents the name of the original depositor, the date of deposit, and key scientific information associated with the deposit. Specifically, ATCC lists the original research reference linked to deposited materials. Often, the original article associated with a material is written by the depositor herself, although, in some cases, materials are deposited by researchers engaged in related work. In its catalog of available cell cultures, ATCC lists both an originating article, as well as additional publications associated with each material.[16]

[16] The ATCC scientific and information technology staff report that the first reference article is typically the one most closely associated with the initial use of the biological material. Historically, ATCC published its catalogs in print form. Currently, ATCC maintains its catalog online at www.ATCC.org.

To overcome the second challenge, we take advantage of shocks that lead to the bulk transfer of materials into ATCC from other collections. Specifically, we investigate the impact of the transfer of three *special collections* that were transferred from collections in the peer-to-peer network to ATCC at specific points in time. These materials transfers occurred when scientists who maintained collections within the peer-to-peer network moved or faced an institutional funding limitation (unrelated to the specific collection) that spurred transfer to a BRC. The first set of materials is drawn from the Tumor Immunology Bank (TIB), which was transferred from the Salk Institute in 1981 due to Institute-wide funding pressures and was accessioned beginning in 1982. Seventy-seven articles associated with the TIB collection deposits appear in the dataset. The second set of articles is associated with materials in the Human Tumor Bank (HTB). Researchers at Sloan-Kettering had maintained the HTB until institution-wide funding considerations led to its being transferred into ATCC beginning in 1981. Forty-four articles associated with HTB deposits appear in the dataset. Finally, the third 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 accept a position at UT-Southwestern. The materials in the Gazdar collection were accessioned beginning in 1994 and are linked to six research articles.

It is important to note two characteristics of these special collections prior to their accession by ATCC. First, each of these special collections was publicly available and was maintained by researchers who actively exchanged the materials with colleagues. Second, the inclusion of these materials in the special collections was tantamount to a commitment not to commercialize innovations derived from these materials. If scientists had intended to assert intellectual property rights over these materials, the materials would not have been freely circulated prior to their transfer into ATCC.

It is also important to note that we choose to study these collections specifically because our interviews with researchers at ATCC suggested that their accession occurred for reasons that were unrelated to changes in the diffusion of knowledge associated with their materials.[17] Specifically, our interviewees suggested that the funding constraints that precipitated the

[17] We are grateful to Dr. Raymond Cypess, President and CEO of the ATCC, and Dr. Robert Hay, Director of the Department of Cell Biology at ATCC, as well as other ATCC staff members, in particular for discussions on the special collections and the history of ATCC.

accession of the TIB and HTB collections were institution-wide and were not related to particular characteristics of these collections or their increasing use by scientists in the few years directly prior to accession. As well, the moves of Dr. Gazdar and Dr. Minna from the NCI to UT-Southwestern appear to have been motivated by professional considerations unrelated to changes in the perceived importance and use of the materials in the special collection they had maintained.

That said, we are also able to take advantage of difference-in-differences techniques to evaluate whether the data suggest that, indeed, the accession of these materials was not preceded by a significant boost in their use. In particular, since the materials included in each collection are associated with articles which are published at different points in time, and each of the special collections is moved at a given point in time, the articles associated with each collection vary in terms of how much time has elapsed between initial publication and BRC deposit. This allows us to estimate the impact of BRC deposit separately from the impact of article age. By examining the trend in citations to special collections' articles in the years preceding deposit (and finding these to be insignificant, although the post-deposit trends are significant), we are able to conclude with some degree of confidence that the "trigger events" leading to the accession of three collections we study were unrelated to pre-deposit changes in intrinsic value of the knowledge associated with their materials.

To address the third challenge, we match each BRC-affiliated article with two types of control articles. We choose these with the aim of ensuring that the control articles are as similar to the BRC-associated article on as many observable dimensions as possible in order to ensure that differences in citation rates will reflect the impact of article-specific differences on knowledge diffusion. The first set of controls is composed of the set of research articles that immediately precedes the article associated with each ATCC deposit in the journal in which the ATCC-linked article was published (we refer to these as the *Nearest Neighbor* controls).[18] For example, if an ATCC-associated publication were the third article in the June 14, 1986 issue of *Cell*, our control article would be the second article within that same issue.[19] By matching control articles to treatment articles in this way, we attempt to minimize heterogeneity associated

[18] We identify *Nearest Neighbor* controls for each BRC-affiliated by using the PUBMED database of scientific journals. PUBMED is a database and search engine constructed and maintained by the National Library of Medicine that provides access to article information contained in the MEDLINE database of journal citations and abstracts. A complete description of PUBMED and MEDLINE can be found at www.pubmed.com.

[19] When the ATCC-associated article is the lead article, we use the second article in that issue as the control.

with the publication process. Specifically, this method ensures that both the BRC-affiliated article and the control article have undergone the same type of scientific review process and have been published at the same moment in time. Comparing the citations by future researchers to these articles provides an indication of their relative impact, conditional on these *ex ante* similarities. Our second set of control articles is based on identifying the most-related article in the same volume of the journal that the BRC-linked article was published (we refer to this set as the *Most-Related Article* controls). To accomplish this, we take advantage of an online search algorithm developed by the National Library of Medicine (NLM) that allows PUBMED to identify a set of articles that mostly closely resembles a selected article and rank them according to similarity. This algorithm determines similarity rankings based on the extent to which articles share terms in their title, abstract, and Medical Subject Headings (MeSH).[20] From the set of articles identified by NLM as related to the focal article, we select the most related article published in the same publication year.[21]

Each of the two control groups provides a useful comparison to the BRC-linked articles. The *Nearest Neighbor* method minimizes the heterogeneity associated with the publication process and eliminates heterogeneity associated with publication timing; the *Most-Related Article* accounts for field-specific within-journal heterogeneity. This second type of control will be particularly important for more general-interest journals (e.g., *Nature* versus the *Journal of Cell Biology*). By including this second control group, we can account for differences in citation patterns in a way that is independent of field-specific norms.

To address the fourth challenge (accounting for ambiguity in the date at which BRC deposits are available for access by other researchers), our dataset accommodates institutional aspects of the accession process. On the one hand, prior to the date of formal accession, the research community becomes informed about collections transfer through formal announcements and informal communications. As a result, materials that are deposited are often known to be part of the transfer prior to the official accession date. On the other hand, because of the rigorous procedures used to accession materials (and short-term limitations on the supply of some materials), accessioned materials are sometimes not made fully available to the research

[20] Medical Subject Headings (MeSH) headings are subject headings developed by the National Library of Medicine to help index articles in the life sciences. They are similar in function to Journal of Economic Literature classifications. A more complete description of the NLM matching algorithm appears at: http://www.ncbi.nlm.nih.gov/entrez/query/static/computation.html.

[21] In some cases, no article in the same volume of the journal qualifies as sufficiently related according to the NLM algorithm. In these instances, we rely on the "Nearest Article" control.

community until many months after the official accession date. In some cases, materials in the HTB and TIB collections took up to 24 months to be declared officially available from ATCC. We explicitly account for the impact of this in our empirical analysis by incorporating a "transfer window," including the year before, the year of, and the year following the official accession date. By including this window, our analysis focuses on how the pattern of citation changes from a period prior to the deposit announcement and subsequent to its availability through a BRC.

Having assembled this dataset of treatment and control articles, we compile additional article-specific data and tabulate annual citation counts from the Science Citation Index Expanded (SCI). The Science Citation Index is a database maintained by the Institute for Scientific Information (ISI) that records reference information for nearly six thousand scientific and technical journals in approximately 150 disciplines. The SCI has been widely used in economics, sociology, and management research, as well as in bibliometric studies, to quantify scientists' research output, measure research collaboration, and track the diffusion of scientific knowledge – prominent examples in economics include Levin and Stephan [1991]; Adams and Griliches [1998]; Henderson and Cockburn [1998]; and Zucker and Darby [1998].

V.B. Summary Statistics

Table I provides variable names and definitions and Table II reports summary statistics. The complete dataset contains the special collections sub-samples and the two sets of associated control articles. For each article 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 289, and the total number of article-year observations is 6475. The overall distribution of ages of articles in the sample is reflected in Figure I. This distribution is centered around 1981-1982, which are the years in which the TIB and HTB collections, respectively, entered into the ATCC.

The key dependent variable in our analysis is FORWARD CITATIONS, the number of articles that reference the focal article in a given year. The average level of citations received by articles in this dataset is 6.23, which is quite high relative to the average among all academic articles. In part, this occurs because the publications associated with BRC deposits (and their associated control articles) tend to appear in top-tier journals, such as *Science, Nature,* and *Cell.* Consistent with most citation analysis, the distribution of citation counts is quite skewed (Figure

II). By the end of 2001, the average article in our sample has received more than 79 total citations.

Key control variables in the analysis are the calendar YEAR, which ranges from 1970 to 2001, and AGE, which equals 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. We also track PRICE for each of the materials in the special collections; this averages approximately $233 per material.

While our analysis focuses mostly on specifications that address article heterogeneity by including article fixed effects, we have collected characteristics about each of the articles in our sample. Specifically, we have assembled information on the number of pages for each article (# PAGES), the number of authors (# AUTHORS), and the number of backward citations (BACKWARD CITATIONS). Although SCI data do not make it possible for all articles, we record whenever possible whether the lead author is associated with a university (UNIVERSITY) or government institution (GOVERNMENT) and whether their institution is located in the United States or another country (NON-US). University researchers comprise the majority of lead authors in the sample (51%); authors affiliated with a government agency comprise 18% of lead authors. The vast majority of lead authors are from U.S. institutions; 29% of authors are from institutions outside of the United States. For our extended analyses, we also compute the fraction of papers associated with lead authors from a set of top university biology departments (TOP UNIVERSITY).

V.C. Comparing citations to BRC-associated articles versus control group articles.

Table III compares the BRC-linked articles to the control groups. Strikingly, articles associated with BRC deposits receive significantly more citations than matched control articles. BRC-associated articles receive, on average, more than four times as many citations as *Nearest Neighbor* controls, and more than 260% more citations than *Most-Related Article* controls. These substantial differences in overall citation exist, even though both control groups appear in the same journal, went through the same review process, and (particularly in the case of the *Most-Related Article* controls) are matched closely in terms of subject area.

Figures III-1 and III-2 portray the disparity between these groups over time, comparing average citations by article age for each control group. Figure III-1 compares citation levels. For each control group, the number of citations increases over the first few years, peaking around the third or fourth year after publication, and deteriorating at various rates over time. In each of the first twenty years after publication (excepting for the publication year, in which all articles receive few citations), the average BRC-associated article receives substantially more citations than control group articles. Moreover, Figure III-2 demonstrates that the "citation premium" received by BRC-associated articles persists or increases, as a percentage of citations, over the first twenty years after an article's publication.

These conditional means suggest that strong differences exist between BRC-linked articles and those in the control groups. While the differences in the citation rates for BRC vs. control articles are substantial and are of primary interest in the study, it is interesting to note that important differences exist between the citation counts of *most-related article* vs. *nearest neighbor* controls. The citation pattern for BRC articles is more similar to the *most-related article* controls than those of *nearest neighbor* controls. In our analysis, we check the robustness of our results to including each control group separately.

VI. EMPIRICAL RESULTS

Our empirical approach relies on a differences-in-differences analysis that separately identifies selection effects from the marginal impact of ATCC deposit. This strategy relies on observing BRC-linked articles in two distinct institutional environments, associated with a pre-deposit and post-deposit period. By comparing citation patterns *across article families* (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 precisely identify the marginal impact of BRC deposit on the rate of knowledge diffusion. Specifically, after controlling for other factors, a positive and significant effect on BRC ARTICLE implies a selection effect (this coefficient reflects the additional citations received by the set of articles that are ultimately accessioned into the BRC), while a positive and significant effect on BRC ARTICLE, POST-DEPOSIT indicates the estimated marginal impact of BRCs (this coefficient measures the "boost" in citations received by BRC-accessioned articles in the period following their deposit, controlling for the fact that they were articles that were, ultimately, accessioned).

It is useful to recall that we incorporate a three-year "window" for the period of time between the announcement that materials will be accessioned by ATCC and the time when they are readily available (and we also check the robustness of our results to the inclusion or exclusion of data within the window period). Over the specifications, the results are consistent with statistically and economically significant evidence for selection into and the marginal impact of BRC deposit.

VI.A. Baseline Analysis

Our analysis begins in Table IV, where we begin to distinguish the selection effect from the institutional impact of BRCs. Recall that specifications that include article family effects allow us to separately identify both the selection and marginal effects. Equations (4-1) and (4-2) present OLS specifications with ln(FORWARD CITATIONS) as the dependent variable. The specifications differ in that (4-1) includes AGE fixed effects, while (4-2) also includes Article Family fixed effects as well as Year fixed effects. The results are similar. In (4-1), the coefficients on both BRC-ARTICLE and BRC-ARTICLE, POST-DEPOSIT suggest a significant impact of BRC association. On average, articles that are ultimately linked to BRC deposits have a 46.5% higher citation rate (relative to the control articles in their "article families"), and receive an additional 61.0% increase in their citation rate after BRC deposit. BRC-linked articles also experience a citation boost during the WINDOW PERIOD, although the boost associated with the window period (35.1%) is significantly smaller than that experienced in the years after accession. These results suggest that both selection and accession effects impact citation rates in a statistically significant and economically important way: BRC-linked articles are cited more frequently and receive an additional "boost" in the years following BRC deposit. In these and in all subsequent specifications, year and article age fixed effects are each jointly significant (though the interpretation of year and age fixed effect coefficients in this context is subtle [Hall et al, 2005]).

Though useful as a preliminary exercise, OLS is inappropriate for inference as citation data are composed of highly skewed count data. We therefore employ a conditional fixed effects negative binomial specification in the remainder of the analysis. We report in brackets the coefficients for these models as incidence-rate ratios (a coefficient equal to one implies no effect on FORWARD CITATIONS, whereas a coefficient equal to 1.50 implies a 50% boost to

FORWARD CITATIONS). For each of these models, we compute bootstrapped standard errors, clustered either by article families or article dummies, depending on the set of fixed effects included in the specification [MacKinnon, 2002]. (Underneath the bracketed IRRs, we report the non-exponentiated regression coefficients, along with their associated bootstrapped standard errors in parentheses.) We do not report the significance of tests of joint restrictions on the article family or article fixed effects, as these are not computed in conditional fixed effects models.

The first of these specifications (4-3) presents a useful comparison to (4-2), insofar as it includes an identical set of regressors. After accounting for citations as skewed count data, we easily reject the null of no selection and no marginal effect. Indeed, the estimated coefficients are larger than those associated with the OLS specifications (e.g., citation rates are estimated to increase more than 90% *after* BRC deposit). In (4-4), the second of these specifications, we disaggregate the selection effect by special collection, replacing BRC-ARTICLE with dummy variables for each of the three special collections, while retaining the common BRC-ARTICLE, POST-DEPOSIT dummy. The results demonstrate that the selection effect is significant and economically substantial for each collection and suggest that the selection effect is greatest for the Gazdar collection.

Overall, the findings in (4-1) through (4-4) provide evidence of both a selection effect and a marginal impact of BRCs on the diffusion of scientific knowledge. While these specifications separately identify the selection and the marginal impact of BRCs, we have so far ignored the substantial variability among articles, even within article families. As a consequence, we introduce a number of specifications that include article fixed effects. Although these do not identify the average selection effect, they do implement a more precise control structure for the impact of individual articles. In these specifications, the coefficient on BRC-ARTICLE, POST-DEPOSIT reflects the "boost" in citation rate that an article receives after its key material is accessioned (and after the deposit window has elapsed).[22] As demonstrated in (4-5), the average article is estimated to experience a 135.0% citation boost after BRC accession (and a 61.2% boost in citations during the WINDOW PERIOD), even after controlling for all article, age and year-specific effects. This suggests that, even controlling for

[22] In these specifications, the pre-deposit period for BRC-accessioned articles effectively serves as the control for identifying the post-deposit impact on FORWARD CITATIONS. The *Most-Related Article* and *Nearest Neighbor Articles* are useful in these specifications for helping to establish year and article age effects.

the overall impact of an article over time, BRC-associated articles experience an economically substantial increase in knowledge diffusion (controlling for year and article age). We view this as one of the core findings in our analysis.

VI.B. Robustness to Alternative Control Groups and Specifications

Table V explores the robustness of the analysis in Table IV to alternative control groups and specifications. Employing the same article fixed effects approach as in (4-5), (5-1) finds that the impact of deposit varies by collection. Articles associated with the HTB and TIB collections experience post-deposit citations boosts of 182% and 128%, respectively, while the articles associated with the Gazdar collection experience only a 55% post-deposit citations boost. The lower impact of deposit on citations experienced by the Gazdar articles may be an artifact of their being accessioned later (and, thus, having a smaller number of years over which a post-deposit effect could be observed). Equations (5-2) and (5-3) following the basic specification implemented in (4-3), but include only the *nearest neighbor* and *most-related article* controls, respectively. The overall post-deposit effect remains positive in both cases, though the magnitude of the effect is significantly higher when compared to the "most related" control. In (5-4), we simply drop the data associated with the window period. The results are similar to our core model.

VI.C. Exploring Persistence and Timing

Our analysis so far has assumed that BRC-linked articles and the control articles follow a similar time trend. We relax this assumption in Table VI, and explore the impact of a BRC-linked time trend in several ways. In (6-1), we introduce a separate BRC-article time trend to account for the possibility that articles associated with BRC deposits may follow a different trajectory with respect to the timing of their citations. BRC-ARTICLE*TIME TREND does enter positively and significantly, suggesting that the citation rate to BRC-associated articles increases over time. Nonetheless, BRC-ARTICLE, POST-DEPOSIT remains positive, significant, and of an important economic magnitude (46.0%). This suggests that, although BRC-affiliated articles are characterized by an additional upward citation trend, these articles also experience a post-deposit citation boost even when controlling for this trend (and year, article age, and article fixed effects).

Simply allowing a BRC-linked time trend is inadequate, however, for two reasons. First, if the time trend is statistically significant in the years prior to deposit, this would cast doubt on

the exogeneity of the timing of deposit (e.g., a third factor was driving both the deposit decision and the increased citation rate). Second, according to the preservation hypothesis, the impact of BRC deposit should increase over time, and so the presence of a post-deposit positive time trend actually provides additional evidence of the marginal impact of BRCs on knowledge diffusion. In other words, rather than simply needing to demonstrate robustness to a BRC-linked time trend, we need to evaluate the pre-deposit and post-deposit trend separately.

We implement this idea in (6-2) and Figure IV. Though positive, the pre-deposit time trend is insignificant. In contrast, the post-deposit time trend is positive and significant (and the coefficient on BRC-ARTICLE, POST-DEPOSIT is statistically and quantitatively significant). According to (6-2), the post-deposit citation boost increases by 3.7% in each year that elapses after the deposit date. To explore these ideas in greater detail, Figure IV presents a specification similar to Equation (3), but with separate dummy variables for each year preceding and following BRC-deposit (along with the complete set of article, age, and calendar year fixed effects). Figure IV plots each coefficient (in terms of the incidence-rate ratio minus one), excluding the years associated with the accession window (all effects are computed relative to the window period). Two findings stand out. First, the pre-deposit citation pattern does not suggest a clear upward trend in the nine years prior to accession; however, the third and second years prior to the window period do appear to show a slight uptick in forward citations. This uptick is, however, sensitive to the estimation technique and we cannot reject the hypothesis that the pre-deposit effect in the second year prior to deposit is statistically greater than the effect in the fourth year prior to deposit.[23] Nonetheless, the result suggests some degree of caution in concluding that the special collections deposits are econometrically exogenous. The second finding does, however, breed substantial confidence BRC-accession has a significant marginal impact on FORWARD CITATIONS relative to the pre-deposit period. There is a sizeable and near continuous increase in the citation boost in the years following deposit. While BRC-affiliated articles experience only a 20% citation boost in the years immediately after accession, this effect increases to over 100% by ten years after deposit (and continues to increase from there). In other words, while the immediate impact is positive but modest, the influence of BRC deposit over time becomes larger, consistent with the preservation hypothesis. Whereas most

[23] In particular, this uptick does not appear in models using the fixed effects estimator with robust standard errors, whose use is recommended by recent results about the relative size and importance of the small sample versus asymptotic bias arising in count data models (Allison and Waterman, 2002; Greene, 2004).

research is relevant and used as an input in follow-on research for only a few years following publication, BRC-accession leads research to be "forgotten" at a significantly lower rate.

Overall, the citation evidence is principally consistent with the hypothesis that the Special Collections were not accessioned because the scientific community recognized these materials as increasingly important prior to their deposit and is quite consistent with the hypothesis that BRC-accession accelerated the use of knowledge associated with those materials.

VI.D. Exploring the Drivers of Marginal Impact – Transactions Costs vs. Certification

The results in Tables V through VI document a significant and long-lived marginal impact of BRCs on FORWARD CITATIONS. Our analysis so far, however, has not attempted to identify the characteristics of BRCs that drive these results. At least two sets of factors consistent with Mokyr's framework [2002] could be at work. First, BRC-accession could drive down the costs of obtaining research materials in comparison to prior institutional arrangements (i.e., which could be described as a reduction in transaction costs). Second, BRC-accession could also yield a certification effect, which increases the scientific community's belief in the fidelity of materials associated with BRCs and result in increased use of knowledge associated with deposited materials. Our interviews with biologists suggest the importance of both factors.[24]

Decomposing the marginal impact of BRCs into separate cost- and certification-based components is difficult in the context of a single institution or in the absence of an experiment that affects these factors in different ways. Although our analysis is limited to one institution, we can take advantage of the characteristics of our experiment and data to obtain suggestive evidence regarding the presence of cost-reduction and certification effects. The key aspect of the experiment that we can exploit is that the movement of the special collections from their prior host institutions into ATCC involves a *shift* in the costs of accessing those materials and in the level of certification associated with each material. Focusing on changes in access costs, two pieces of evidence can suggest the presence of cost-reduction effects. First, we know the PRICE at which each of the materials in the special collections is available from ATCC. If access costs decrease when materials are shifted into ATCC, we would expect the impact of such cost reductions to be less for materials with higher PRICE. Thus, we would anticipate the incidence

[24] For example, Arnold Demain of MIT's Department of Biology, explained that "it would probably be better if all (or, at least, very many) collections were moved to ATCC, because people don't have the time or training to properly maintain their own collections." Thus, he continued, scientists can place greater trust in materials they obtain from ATCC than from scientists directly. He also noted in our conversation, however, that obtaining materials through ATCC was often easier than dealing with scientists directly, particularly as universities become increasingly concerned with intellectual property rights.

rate ratio for PRICE to be below unity. Second, we can take advantage of the fact that, while the changes in access costs should be the same within collections (controlling for PRICE), they will likely differ systematically across collections, as each of the special collections had different access costs prior to deposit but end up with equal access costs after BRC-deposit. This fact also helps us identify certification effects: Because the changes in access costs are constant within collection (when controlling for PRICE), systematic within-collection differences in the way in which deposit impacts follow-on research will reflect certification effects rather than changes in access costs. In particular, we hypothesize that the certification boost will be greater for individuals that were less prominent prior to deposit and for articles that were less prominent prior to deposit.

We investigate these mechanisms in Table VII.[25] Equation (7-1) provides evidence that, indeed, access costs affect the extent of follow-on research. PRICE * POST-DEPOSIT is negative, statistically significant, and of an economically significant magnitude for each collection, suggesting that materials for which access costs are relatively higher (i.e., those materials available at a PRICE greater than their collection average) are associated with a lower impact of BRC deposit.[26]

We investigate certification effects by evaluating whether variations in article and author prominence before deposit have an impact on post-deposit citations. To do so, we must control for factors that affect the cost of access to materials, including PRICE and the identity of the special collection with which the deposit is associated. Controlling for these factors, the estimates in (7-2) suggests that NON-US articles do receive an additional boost in citations relative to articles with US-based authors. (NON-US * HTB-ARTICLE * POST-DEPOSIT enters positively and significantly in each equation, and a test of joint restrictions finds NON-US * HTB-ARTICLE * POST-DEPOSIT and NON-US * TIB-ARTICLE * POST-DEPOSIT to be jointly significant at the 1% level.)[27] We also test whether a articles in a collection that were less prominent prior to deposit (greater than collection median citations) or that were authored by researchers from less prominent institutions (not TOP UNIVERSITY) experienced a greater citation boost after accession than other articles in their collection. The coefficients on these variables are not significant, however, and are not reported in Table VII. The result described

[25] Note that the analysis in Table 7 does not include the Gazdar collection, because we were unable to determine a sufficiently large number of Gazdar article characteristics with sufficient certainty (though the results are robust to the inclusion of Gazdar data of which we are confident).

[26] PRICE has been demeaned in the analysis at the level of the Special Collection, so that values of PRICE reflect variation from the average price of all materials in its collection.

above that the within-collection citation boost is higher for NON-US authors is, however, robust across specifications. Overall, we interpret these results as providing suggestive evidence that both access costs and certification effects contribute to the impact of BRC deposit on the exploitation of scientific knowledge and that these can be identified separately.

VII. DISCUSSION

While growth theorists, industrial organization economists, and economic historians have each come to place increasing importance on the role of cumulativeness in sustaining innovation, little research has directly addressed the microeconomic conditions supporting cumulativeness, or provided direct statistical evidence about the impact of institutions in enhancing the cumulative knowledge production process. In this paper, we investigate the role of institutions in this process directly. Specifically, we consider research-enhancing institutions, which facilitate step-by-step scientific and technical progress by leveraging the potential of research from one generation to serve as "seed corn" for future generations.

Our principal contribution has been to provide evidence about the role of BRCs as a specific institution within the life sciences that impact the cumulativeness of research in that field. We highlight a class of institutions whose role in the advance of the life sciences appears to be vital and study the specific importance of one of these institutions in cumulative knowledge diffusion. By being precise about the institution we study and by focusing on specific episodes of deposits to this institution, we are able to develop and implement a novel methodology for evaluating how institutions matter in the context of citation analysis and are able to provide the direct evidence regarding the extent to which these institutions matter for the process of cumulative research production. In particular, we separate out a selection effect (which turned out to be quite important) from the marginal impact of the institution in enhancing knowledge diffusion, and identify suggestive evidence that both reduced access costs and certification benefits play a role in driving the marginal impact of BRCs on follow-on research. Each of these effects is significant for understanding the role of this institution. On one hand, the BRC we study plays an important role in the process of knowledge cumulation in life sciences by identifying and preserve research materials that are, themselves, important to biological research. At the same time, BRCs enhance the value of the materials they collect, thus facilitating the

[27] Note that we cannot conduct this test for the Gazdar collection, since each of its depositors is US-based.

process of standing on the shoulders of prior research. Our evidence suggests that this occurs both because BRCs lower the transactions costs associated with obtaining important research materials and because accessioned materials, especially those of non-US-based researchers, appear certified for potential follow-on researchers. Our results are certainly subject to the caveats associated with any research premised on the use of citation data; however, our estimates do suggest that BRCs provide precisely the type of boost that is at the cornerstone of the economics of cumulative knowledge production.

Our framework and results point to several areas for future research. First, while we present suggestive evidence about the mechanisms underlying the citations boost associated with BRC-deposit, future research could exploit differences between materials in their level of certification and differences in the level of certification across different BRCs as sources of variation to further disentangle these two different functions of BRCs. Second, whereas our econometric exercise exploits specific instances associated with plausibly exogenous deposits, the economics of research-enhancing institutions depends, by and large, on the endogenous decision by individual researchers to allow their knowledge to become accessible through research-enhancing institutions. Mukherjee and Stern [2004], for example, address this issue in the context of a simple overlapping generations model that identifies the key factors underlying this endogenous choice, deriving conditions under which investments in institutions that support disclosure and cumulativeness provide a social welfare benefit. Additional research in this vein may focus in a useful way on the incentives for individual scientists and scientific fields to develop research-enhancing institutions as well as study additional such institutions.

In addition, it is important to emphasize that the choice to deposit materials in BRCs (or disclose knowledge through other research-enhancing institutions) is sensitive to parameters that are themselves influenced by public policy. Stern (2004) undertakes a thorough policy analysis of the specific issues relating to BRCs *per se*. However, the insights associated with this study are more general. For example, most policy debates regarding federal research investments focus on expanding the level of research conducted; in contrast, this line of research raises the point that it may be optimal to shift funds towards institutions and other mechanisms to ensure that knowledge, once produced with public funds, is made accessible to future research efforts.

Finally, our empirical approach highlights an important but often overlooked problem in the measurement of knowledge spillovers. Simply put, it is difficult to disentangle the impact of

institutions from the knowledge that is accessible from those institutions. While sharp insights have been developed over the past decade about the relationship between university research and follow-on commercialization, these prior studies have not been able to clarify whether the "boost" associated with university research is the result of differences in the type of research conducted or the rules and policies governing the disclosure and dissemination of university research results. However, these rules are precisely what is at issue in terms of contemporary policy discussions [Heller and Eisenberg, 1998; Argyres and Liebeskind, 1998; David, 2001; Murray, 2002]. Murray and Stern [2005] appears to be a useful effort at applying the techniques we develop here to address that particular question. More broadly, however, we believe the careful exploitation of experiments and difference-in-differences techniques constitutes a fruitful and as yet underexploited methodology for investigating the impact of institutions on knowledge flows.

REFERENCES

Acemoglu, D., S. Johnson, and J.A. Robinson (2002) "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics* 117: 1231-1294.

Adams, J. D. and Z. Griliches (1998) "Research Productivity in a System of Universities," *Annals of INSEE*, pp. 127-162.

Allison, P. D. and R. Waterman (2002) "Fixed effects negative binomial regression models." Ross Stolzenberg (ed.), *Sociological Methodology 2002*. Boston: Basil Blackwell.

Argyres, N. and J. Liebeskind (1998) "Privatizing the Intellectual Commons: Universities and the Commercialization of Biotechnology Research," *Journal of Economic Behavior and Organization.*

Arrow, K. (1962) "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity*, edited by Richard R. Nelson, 609-25.

Bertrand, M, E. Duflo, and S. Mullainathan (2004) "How Much Should We Trust Difference in Differences Estimates?" *Quarterly Journal of Economics* 119(1): 249-275.

Branstetter, L. (2003) "Exploring the Link Between Academic Science and Industrial Innovation: The Case of California's Research Universities," mimeo, Columbia University.

Branstetter, L. and M. Sakakibara (2002) "When Do Research Consortia Work Well and Why? Evidence from Japanese Panel Data," *American Economic Review*, 92(1): 143–159.

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

Cypess, R. (2003) *Biological Resource Centers: From Concept to Reality*, American Type Culture Collection.

David, P. (2001) "From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science'" Stanford Working Paper #01-006.

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

Hunter-Cevera, J.C. (1996) "The importance of culture collections in industrial microbiology and biotechnology, in Culture Collections to Improve the Quality of Life," eds. R. A. Samson, J. A. Stalpers, D. van der Mei et al., Centraalbureau voor Schimmelcultures, Baarn (The Netherlands), 158-62.

Gallini, N. and S. Scotchmer, (2002) "Intellectual Property: When is it the Best Incentive Mechanism?," *Innovation Policy and the Economy*, Vol 2, Adam Jaffe, Joshua Lerner and Scott Stern, eds., MIT Press, pp. 51-78.

Glaeser, E.L., S. Johnson and A. Shleifer (2001) "Coase Versus the Coasians," *Quarterly Journal of Economics* 116: 853-899.

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

Greene, William. 2004. "Fixed Effects and the Incidental Parameters Problem in the Tobit Model," *Econometric Reviews.*

Griliches, Z. (1990) "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 92: 630-653.

Griliches, Z. (1998) *R&D and Productivity: The Econometric Evidence*, University of Chicago Press.

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

Hausman, J, B. Hall, and Z. Griliches (1984) "Econometric Models for Count Data with an Application to the Patents R&D Relationship," *Econometrica*, 52: 909-938.

Hall, B., J. Mairesse, and L. Turner (2006) "Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists," *Economics of Innovation and New Technology*, forthcoming.

Heller, M. and R. Eisenberg. (1998) "Can patents deter innovation? The anticommons in biomedical research," *Science*, vol. 280.

Henderson, R.M. and I.M. Cockburn (1998) "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery," *Journal of Industrial Economics*, 46(2): 157-182.

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*, pp. 577-598.

Jones, B. (2003) "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" mimeo, MIT.

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

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

Lerner, J. and J. Tirole (2003) "Efficient Patent Pools" *American Economic Review*, forthcoming.

Levin, S. G. and P. E. Stephan (1991) "Research Productivity over the life cycle: Evidence for academic scientists," *American Economic Review*, 81(1): 114-132.

MacKinnon, J. G. (2002) "Bootstrap inference in econometrics," *Canadian Journal of Economics*, 35(4): 615-645.

Masters, J. R. et al (2001) "Short tandem repeat profiling provides an international reference standard for human cell lines," *Proceedings of the National Academy of* Science. 98(14): 8012-8017.

Masters, J. R. (2002) "HeLa cells 50 years on: the good, the bad and the ugly," *Nature Reviews Cancer* 2: 315-319.

Merton, R. (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. Edited by Norman Storer. Chicago: University of Chicago Press.

Mokyr, J. (2002) *The Gifts of Athena*. Princeton, NJ: Princeton 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.

Mukherjee, A. and S. Stern (2004) "Open Science: Disclosure, Research Investment and Cumulative Knowledge Production," mimeo, Kellogg School of Management.

Murray, F. (2002) "Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering," *Research Policy*, 31(8): 1389-1403.

Murray, F. (2005) “Exchange Relationships & Cumulative Innovation: Standing on the Shoulders of the Oncomouse,” Working Paper.

Murray, F. and S. Stern (2005) “Do formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the anti-Commons Hypothesis,” NBER Working Paper #11465.

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.

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,

Pederson, G., et al. (2005) "Conservation, Characterization, and Evaluation of Crop Genetic Resources and Associated Information 2005 Annual Report," United States Department of Agriculture, Agricultural Research Service.

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

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

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

Rosenberg, N. (1979) “Technological interdependence in the American Economy,” *Technology and Culture*, 20: 25-50.

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*, 32(1): 77-100.

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

Smith, D. (2003) “Culture collections over the world,” *International Microbiology*, 6(2), 95-100.

Sorenson, O. and L. Fleming (2004) "Science and the diffusion of knowledge," Research Policy 33, 1615-1634.

Stern, S. (2004) *Biological Resource Centers: Knowledge Hubs for the Life Sciences*, Washington (DC): Brookings Institution Press.

Stocker, S. (2006) “Critical Resources - The Microbe and Cell Superstore,” *NCRR Reporter*, 27(3), National Center for Research Resources – Summer 2003, 27(3), 12-13.

World Federation of Culture Collections – World Data Center for Microorganisms, “The Culture Collection in this world,” WDCM STATISTICS, online data: http://wdcm.nig.ac.jp/statistics.html, accessed September 18, 2006.

Zucker, L. G., M. R. Darby, and M. B. Brewer, (1998) “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises,” *American Economic Review*, 88(1): 290-306.

TABLE I
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	SCI
AGE	Year –Article Publication Year	SCI
ARTICLE CHARACTERISTICS		
BRC ARTICLE	Dummy variable equal to 1 if Article is associated with a material deposited in the biological resource center ATCC (the American Type Culture Collection)	ATCC
BRC ARTICLE, WINDOW PERIOD	Dummy variable equal to 1 if Article is referenced by BRC deposit and YEAR = DEPOSIT YEAR or DEPOSIT YEAR plus or minus + 1	ATCC
BRC ARTICLE, POST DEPOSIT	Dummy variable equal to 1 if Article is referenced by BRC deposit and YEAR > DEPOSIT YEAR + 1 (i.e., deposit has already occurred and DEPOSIT WINDOW PERIOD already passed)	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.	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 LEAD AUTHOR	Dummy variable equal to 1 if lead author is associated with a university; 0 otherwise	SCI; author verification
GOVERNMENT LEAD AUTHOR	Dummy variable equal to 1 if lead author is associated with a government-affiliated institution; 0 otherwise	SCI; author verification
NON-US LEAD AUTHOR	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 LEAD AUTHOR	Dummy variable equal to 1 if lead author is associated with a private institution; 0 otherwise	SCI; author verification
TOP UNIVERSITY	Dummy variable equal to 1 if lead author is associated with a university identified by NSF at one of Top 25 universities federally financed R&D expenditures in the biological sciences	NSF; author verification

TABLE II
MEANS & STANDARD DEVIATIONS

VARIABLE	N	MEAN	STANDARD DEVIATION	MIN	MAX
CITATION-YEAR CHARACTERISTICS					
FORWARD CITATIONS	6475	6.23	14.32	0	186
CUMULATIVE CITATIONS	6475	79.28	1616.51	0	2333
YEAR	6475	1989.79	7.21	1970	2001
AGE	6475	11.26	7.22	0	31
ARTICLE CHARACTERISTICS *(N=289 total articles)*					
TOTAL CITATIONS	289	140.10	238.94	0	2333
PUBLICATION YEAR	289	1979.42	4.58	1970	1992
BRC ARTICLE	289	0.37	0.48	0	1
DEPOSIT YEAR*	108	1983.63	3.47	1981	1994
PRICE*	108	233.12	42.60	167	270
# PAGES	287	7.09	6.96	0	69
# AUTHORS	288	4.71	4.24	0	57
BACKWARD CITATIONS	272	31.33	29. 27	0	401
UNIVERSITY^	236	0.51	0.50	0	1
GOVERNMENT^	236	0.18	0.39	0	1
NON-US^	216	0.29	0.45	0	1
TOP UNIVERSITY^	236	0.19	0.45	0	1

* *DEPOSIT YEAR & PRICE are only meaningful for articles associated with BRC deposits.*
^ Institutional affiliations are not available for some of the publications.

TABLE III
MEANS & STANDARD DEVIATIONS, BY CONTROL GROUP

	Treatment Articles	**Control Articles**	
	Articles Associated with ATCC Deposits	**Nearest Neighbor Control**	**Most-Related Article Control***
#PAPERS	108	108	73
PAPER-YEARS	2418	2415	1642
FORWARD CITATIONS	11.13 (19.64)	2.68 (6.91)	4.25 (10.62)
CUMULATIVE CITATIONS	250.50 (331.00)	60.18 (103.98)	97.37 (146.64)
PUBLICATION YEAR	1979.40 (4.55)	1979.40 (4.55)	1979.48 (4.72)

* *There are fewer Most-Related Control Articles than Treatment Articles, because the NIH algorithm is occasionally unable to identify a "most-related article" in the same year and journal as the Treatment Article.*

TABLE IV
BASELINE SPECIFICATIONS

	OLS Dep Var = ln(FORWARD CITATIONS)*		**CONDITIONAL FIXED EFFECTS NEG BINOMIAL** *[Incidence-Rate Ratios reported in brackets] Non-exponentiated coefficients reported without brackets (Bootstrapped Standard Errors reported in parentheses)* **Dep Var = FORWARD CITATIONS**		
	(4-1) Base Model: BRC Effect with Age FEs only	**(4-2)** Base Model, with Article Family & Year FEs	**(4-3)** Baseline Count Model	**(4-4)** *(4-3)* with Collection Effects	**(4-5)** Core Article Fixed Effects Model
ARTICLE CHARACTERISTICS					
BRC-ARTICLE	**0.465** **(0.143)**	**0.513** **(0.122)**	**[2.079]** **0.732** **(0.174)**		
BRC-ARTICLE, WINDOW PERIOD	**0.351** **(0.122)**	**0.355** **(0.105)**	**[1.408]** **0.342** **(0.157)**	**[1.407]** **0.341** **(0.161)**	**[1.612]** **0.477** **(0.147)**
BRC-ARTICLE, POST-DEPOSIT	**0.610** **(0.161)**	**0.574** **(0.132)**	**[1.913]** **0.649** **(0.190)**	**[1.934]** **0.660** **(0.177)**	**[2.350]** **0.854** **(0.144**
HTB-ARTICLE				**[2.110]** **0.747** **(0.229)**	
TIB-ARTICLE				**[2.010]** **0.698** **(0.188)**	
GAZDAR-ARTICLE				**[2.658]** **0.978** **0.212**	
CONTROL VARIABLES					
Parametric Restrictions					
Article Family FEs = 0		**F-stat 5.9*10^5** **p-value 0.00**	**Included**	**Included**	
Article FEs = 0					**Included**
Age FEs = 0	**F-stat 16.84** **p-value 0.00**	**F-stat 14.77** **p-value 0.00**	**χ^2 1608.38** **p-value 0.00**	**χ^2 1285.09** **p-value 0.00**	**χ^2 1007.92** **p-value 0.00**
Year FEs = 0^		**F-stat 2.81** **p-value 0.00**	**χ^2 110.98** **p-value 0.00**	**χ^2 224.64** **p-value 0.00**	**χ^2 182.89** **p-value 0.00**
Regression Statistics					
R-squared	0.26	0.51			
Log-likelihood			-13610.17	13607.11	-11795.38
# of Observations	6475	6475	6371	6371	6298

* Robust standard errors, adjusted for clustering by article group, are in parentheses.

^ Year FEs included for 1980-2001; 1970-1974 and 1975-1979 grouped.

TABLE V
ROBUSTNESS TO ALTERNATIVE CONTROL GROUPS AND SPECIFICATIONS

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL *[Incidence-Rate Ratios reported in brackets]* *Non-exponentiated coefficients reported without brackets* *(Bootstrapped Standard Errors reported in parentheses)* **Dep Var = FORWARD CITATIONS**			
	(5-1)	**(5-2)**	**(5-3)**	**(5-4)**
	Core FE Model: Results by Collection	Only "Nearest Neighbor" Controls	Only "Most-Related Article" Controls	Omitting Window Period Observations
ARTICLE CHARACTERISTICS				
BRC-ARTICLE, WINDOW PERIOD	**[1.620]** **0.483** **(0.150)**	[1.213] 0.193 (0.128)	**[1.921]** **0.653** **(0.220)**	
BRC-ARTICLE, POST-DEPOSIT		**[1.828]** **0.603** **(0.138)**	**[2.297]** **0.831** **(0.228)**	**[2.550]** **0.936** **(0.143)**
HTB-ARTICLE, POST-DEPOSIT	**[2.825]** **1.039** **(0.181)**			
TIB-ARTICLE, POST-DEPOSIT	**[2.285]** **0.826** **(0.196)**			
GAZDAR-ARTICLE, POST-DEPOSIT	**[1.553]** **0.440** **(0.196)**			
CONTROL VARIABLES				
Parametric Restrictions				
Article Family FEs = 0		**Included**	**Included**	
Article FEs = 0	**Included**			**Included**
Age FEs = 0 *(#restrictions = 30)*	**χ^2 1133.23** **p-value 0.00**	**χ^2 718.57** **p-value 0.00**	**χ^2 900.15** **p-value 0.00**	**χ^2 665.92** **p-value 0.00**
Year FEs = 0^ *(#restrictions = 23)*	**F-stat 111.18** **p-value 0.00**	**F-stat 107.72** **p-value 0.00**	**F-stat 102.85** **p-value 0.00**	**χ^2 171.98** **p-value 0.00**
Regression Statistics				
Log-likelihood	-11786.54	-9075.71	-8581.88	-10865.70
# of Observations	6298	4656	3984	5997

^ Year FEs included for 1980-2001; 1970-1974 and 1975-1979 grouped.

TABLE VI
EXPLORING PERSISTENCE AND TIMING

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL ***[Incidence-Rate Ratios reported in brackets]*** ***Non-exponentiated coefficients reported without brackets*** ***(Bootstrapped Standard Errors reported in parentheses)*** **Dep Var = FORWARD CITATIONS**	
	(6-1) Including a BRC-article Time Trend	**(6-2)** Including Pre- and Post-Deposit Trend
ARTICLE CHARACTERISTICS		
BRC-ARTICLE, WINDOW PERIOD	**[1.292]** **0.256** **(0.182)**	**[1.633]** **0.490** **(0.205)**
BRC-ARTICLE, POST-DEPOSIT	**[1.460]** **0.379** **(0.223)**	**[1.649]** **0.500** **(0.288)**
BRC-ARTICLE * TIME TREND	**[1.045]** **0.044** **(0.013)**	
BRC-ARTICLE * PRE-DEPOSIT TREND		[1.060] 0.058 (0.040)
BRC-ARTICLE * POST-DEPOSIT TREND		**[1.037]** **0.036** **(0.015)**
CONTROL VARIABLES		
Article FEs = 0 *(#restrictions = 287)*	**Included**	**Included**
Age FEs = 0 *(#restrictions = 30)*	χ^2 **647.34** **p-value 0.00**	χ^2 **782.66** **p-value 0.00**
Year FEs = 0^ *(#restrictions = 23)*	χ^2 **345.39** **p-value 0.00**	χ^2 **137.89** **p-value 0.00**
Regression Statistics		
Log-likelihood	-11755.96	-11764.01
P-value of Chi	0.00	0.00
# of Observations	6298	6298

^ Year FEs included for 1980-2001; 1970-1974 and 1975-1979 grouped.

TABLE VII
EXPLORING ACCESS COSTS & CERTIFICATION EFFECTS

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL *[Incidence-Rate Ratios reported in brackets]* *Non-exponentiated coefficients reported without brackets* *(Bootstrapped Standard Errors reported in parentheses)* **Dep Var = FORWARD CITATIONS**		
	(7-1)	**(7-2)**	**(7-3)**
ARTICLE CHARACTERISTICS			
BRC-ARTICLE * WINDOW PERIOD	**[1.402]** **0.338** **(0.176)**	**[1.429]** **0.357** **(0.146)**	**[1.409]** **0.343** **(0.172)**
BRC-ARTICLE * POST-DEPOSIT	**[2.207]** **0.792** **(0.180)**	**[2.170]** **0.775** **(0.158)**	**[2.136]** **0.759** **(0.182)**
TIB-ARTICLE * POST-DEPOSIT * PRICE°	***[0.996]*** ***-0.0038*** ***(0.0021)***		***[0.997]*** ***-0.0034*** ***(0.0019)***
HTB-ARTICLE * POST-DEPOSIT * PRICE°	**[0.992]** **-0.0083** **(0.0041)**		***[0.993]*** ***-0.0073*** ***(0.0042)***
TIB-ARTICLE * POST-DEPOSIT * NON-US		[1.298] 0.261 (0.184)	[1.177] 0.163 (0.219)
HTB ARTICLE * POST-DEPOSIT * NON-US		**[2.239]** **0.806** **0.131**	**[1.828]** **0.603** **(0.162)**
CONTROL VARIABLES			
Article FEs	**Included**	**Included**	**Included**
Age FEs = 0 *(#restrictions = 30)*	**χ^2 916.10** **p-value0.01**	**χ^2 1614.37** **p-value0.00**	**χ^2 1365.70** **p-value0.00**
Year FEs = 0^ *(#restrictions = 23)*	**χ^2 60.43** **p-value0.00**	**χ^2 137.48** **p-value0.00**	**χ^2 148.96** **p-value0.00**
Regression Statistics			
Observations	4690	4690	4690
Log Likelihood	-9179.36	-9188.76	-9174.07

° *Note: PRICE has been demeaned in the analysis at the level of the Special Collection, so that values of PRICE reflect variation from the average price of all materials in its collection.*

^ *Year FEs included for 1980-2001; 1970-1974 and 1975-1979 grouped.*

FIGURE I
NUMBER OF PUBLICATIONS BY YEAR

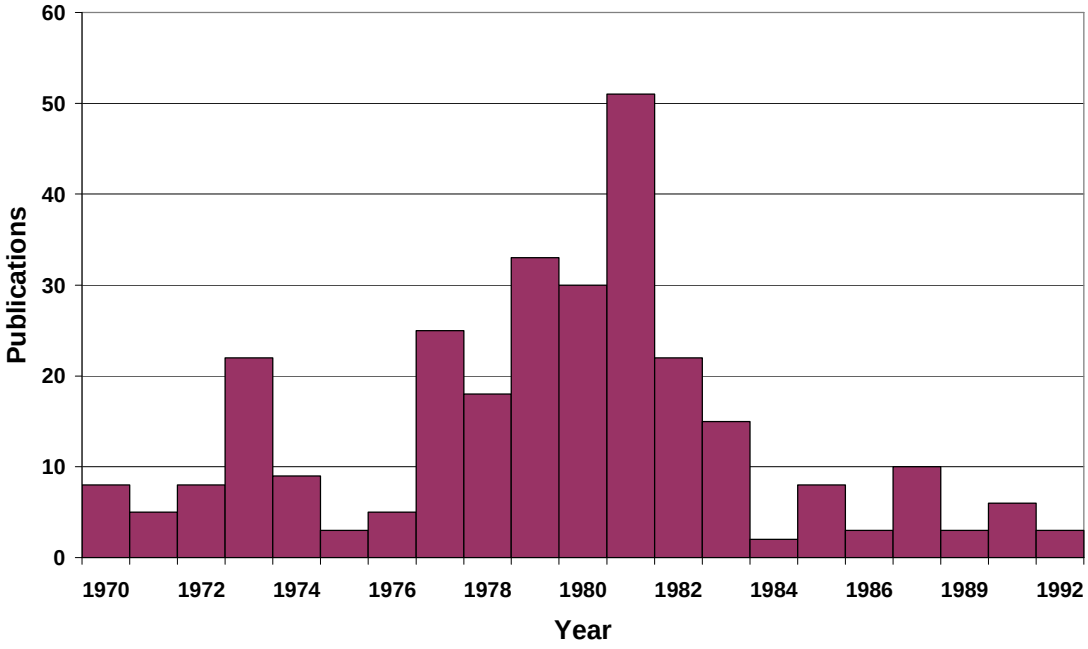


FIGURE II
DISTRIBUTION OF ANNUAL FORWARD CITATIONS

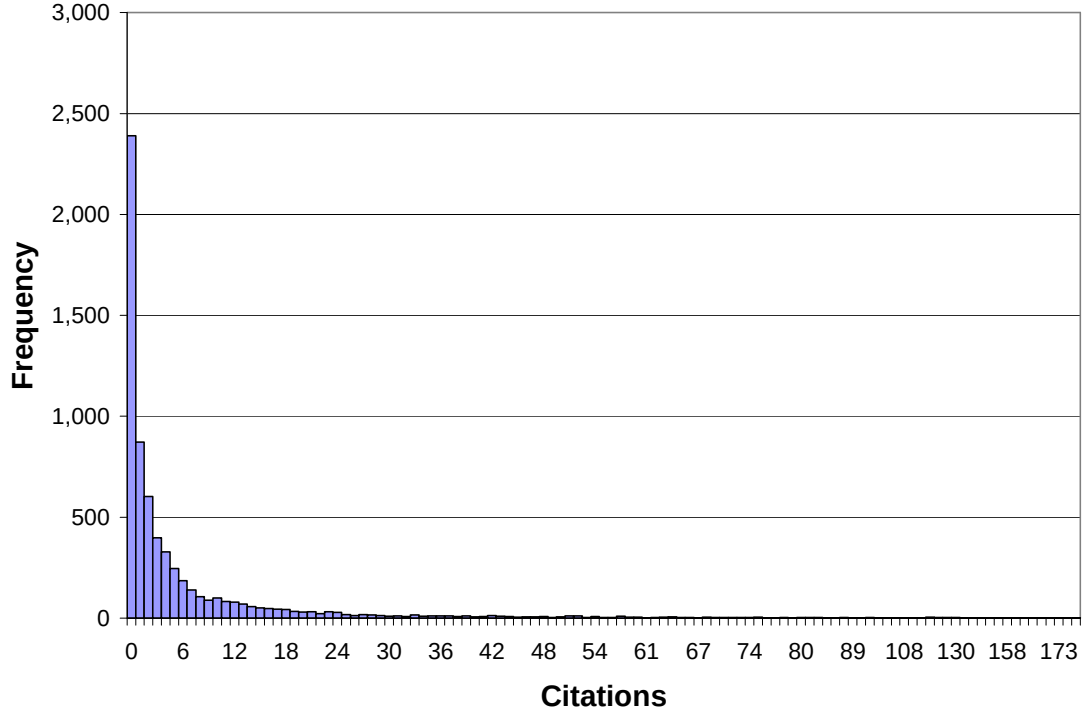


FIGURE III-1
AVERAGE ANNUAL CITATIONS BY AGE, BRC VS. CONTROL ARTICLES

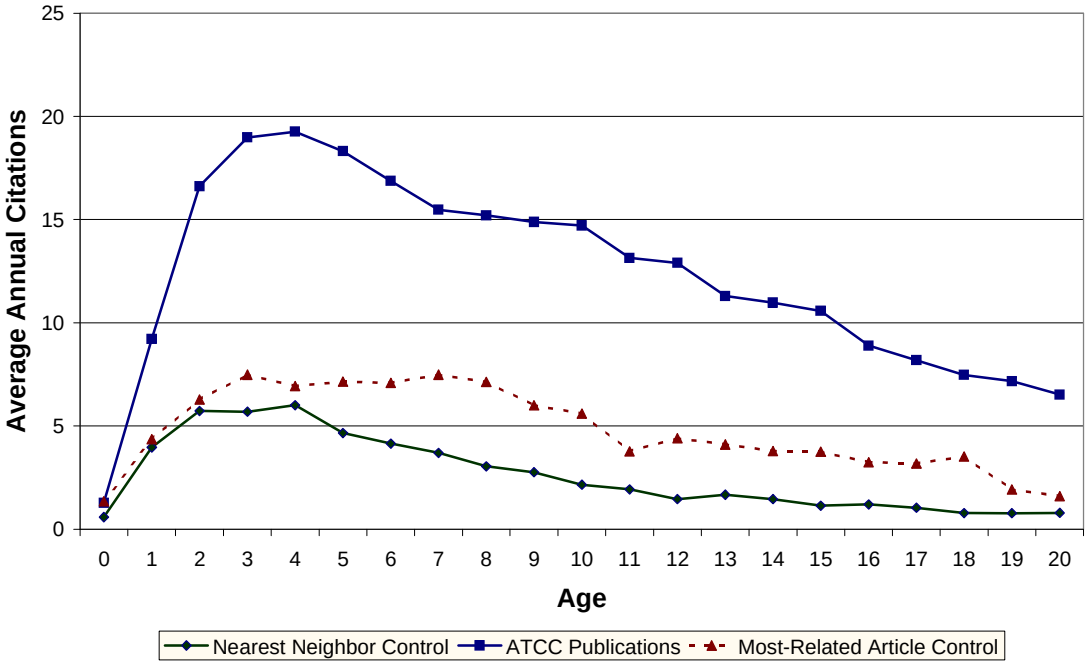


FIGURE III-2
PERCENT DIFFERENCE IN ANNUAL AVERAGE CITATIONS BY AGE, BRC VS. CONTROL ARTICLES

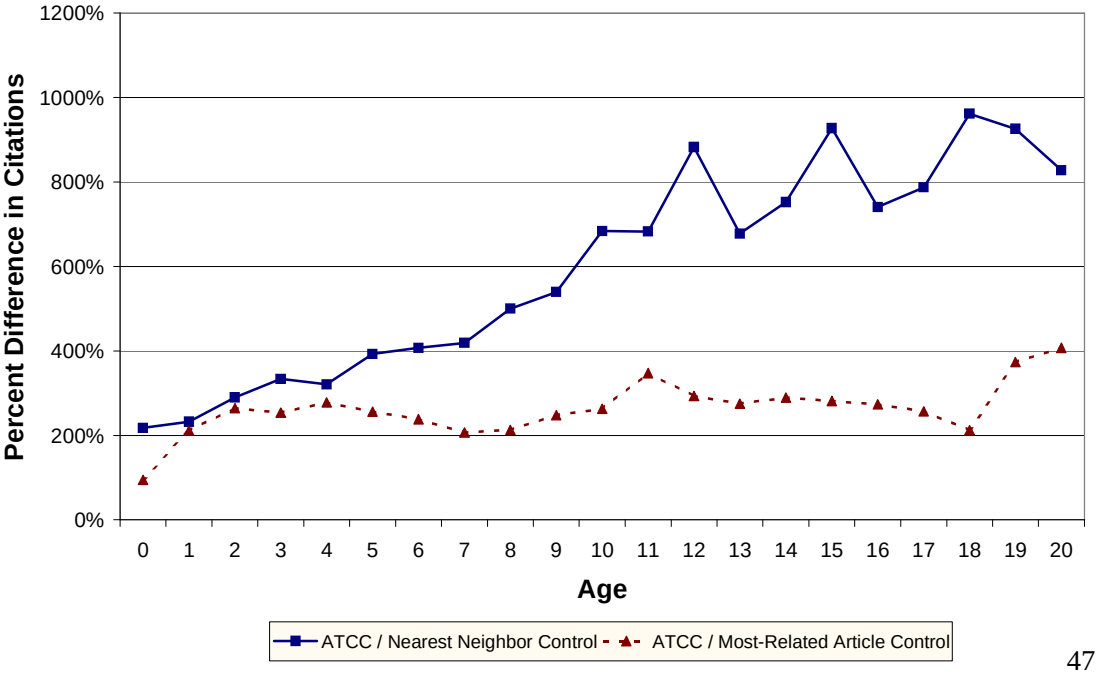
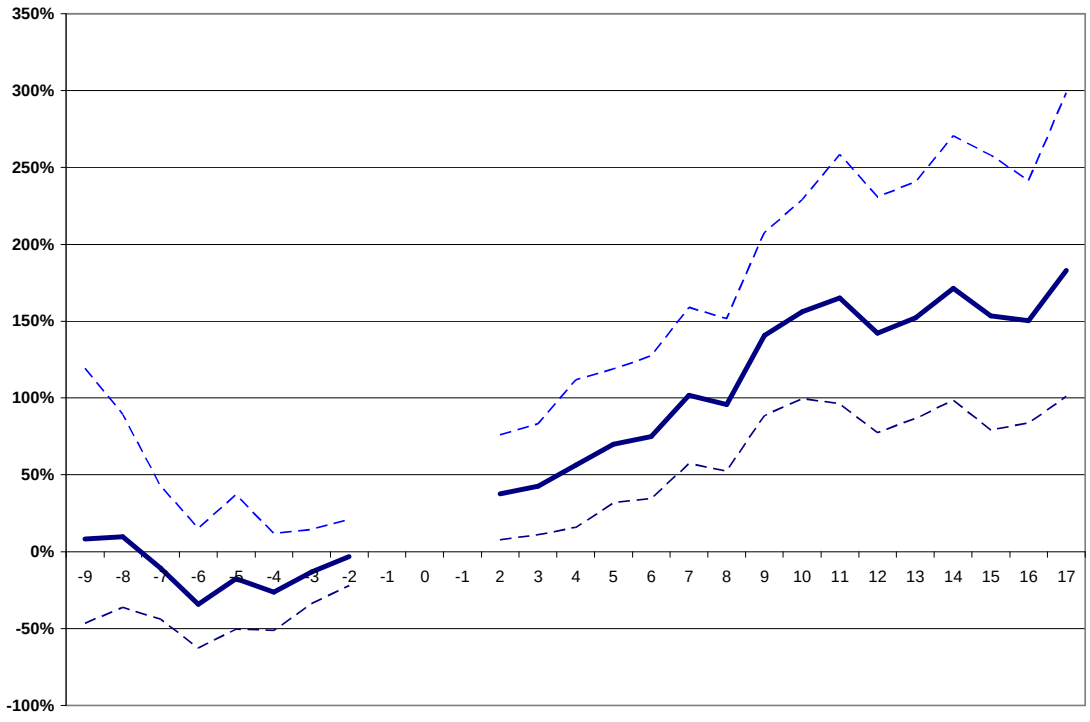


FIGURE IV
PRE- AND POST-DEPOSIT EFFECTS ON FORWARD CITATIONS



APPENDIX TABLE I
TYPES OF BIOLOGICAL RESOURCE CENTERS

Center Type	Examples	
Public / Non-Profit national collections	• ATCC (USA) • DSMZ (Germany) • Japan Collection of Microorganisms (Japan)	Tier 1
Public / Non-Profit specialized collections	• Coriell Medical Research Institute (human genetic mutant cell lines) • National Institute of Allergy and Infectious Disease (HIV materials) • Ribosomal Database Project • Agricultural Research Service Culture Collection (NRRL)	Tier 2
Private, industrial collections	• Merck (antibiotics screening collection, clinical microbiology collection) • Institute for Fermentation Osaka (IFO)	
Specialized University collections	• Yale University (*Escherichia coli* Genetic Stock Center) • Ohio State (*Bacillus* Genetic Stock Center) • Penn State (*Fusarium* Genetic Stock Center)	
Life Sciences Data Management Institutions	• Ribosomal Database Project (RDP) • Institute for Genomic Research (TIGR	

Source: Adapted from OECD, 2001.