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

Jeffrey L. Furman
Boston University

Scott Stern
Northwestern University and NBER

March 2004

* 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, Paul Romer, and participants in numerous seminars. Jason Corradini, Mercedes Delgado, Lorraine DeLeon, Chijoke Emineke, Anna Harrington, Martha Kam, Julia Lo, Kinga Piekos, Megan MacGarvie, Courtney Mason, Tracy Myers, and Naomi Wohl provided excellent research assistance. All errors are our own. Financial support for this research was provided by the Brookings Institution. Author contact information: Scott Stern, Kellogg Graduate School of Management, 2001 Sheridan Road, Evanston, IL 60208 s-stern2@northwestern.edu, and Jeffrey 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

Jeffrey L. Furman
Scott Stern

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 research-enhancing institutions 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. 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 to evaluate how BRCs 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 exogenous shocks. Employing a difference-indifferences estimator linking specific materials deposits to journal articles, we find evidence for both selection into and the marginal impact of BRCs. Moreover, the marginal impact increases with time. Public expenditures towards ensuring that discoveries and knowledge are accessible to future research generations may have a higher rate of return than simply funding additional research studies.

"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. However, 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. The ability of a society to stand on the shoulders of giants depends not only the amount of knowledge it generates, but on the quality of mechanisms for storing knowledge, the fidelity of that knowledge, and the costs to access that knowledge.

Institutions and public policy are often suggested as key to the cumulative process. In assessing the extent to which any institution influences the way in which the "knowledge stock" is created, maintained, and extended, researchers face a considerable challenge. Though conceptually distinct, it is empirically difficult to isolate the impact of a particular piece of

[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).

knowledge from the institution in which it is embedded. 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 the institution on cumulativeness, conditional on the type and quality of knowledge considered. Most prior research conflates the selection effect and the marginal impact of institutions.

The main contribution of this paper is to provide direct statistical evidence for the impact of a specific institution – biological resource centers (BRCs) -- on the cumulativeness of knowledge. BRCs play an important role in life sciences research. BRCs collect, certify and distribute biological organisms for use in biological research and in the development of commercial products in the pharmaceutical, agricultural and biotechnology industries. BRCs maintain large and varied collections of biological materials, including cell lines, micro-organisms, and DNA material, and distribute tools that allow researchers to access and exploit these materials. The ability to exploit prior research in the life sciences depends on access to the cells, cultures, and specimens used in that research. BRCs are a key institutional arrangement by which scientists can obtain materials for research purposes. Our empirical analysis evaluates whether the ability to access research materials through a BRC is associated with enhancing the impact of the scientific research article that initially described those research materials.

At a broad level, our analysis builds on a recent literature that aims to provide quantitative evidence for the role of specific institutions in supporting economic growth (Glaeser et al, 2001; Acemoglu, et al, 2002). 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 two 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 describing its initial characterization and application. Second, various subsets of BRC deposits have been shifted exogenously from prior institutional arrangements into BRCs. For example, some collections maintained in private university laboratories have been shifted into a public BRC when the principal investigator retires or switches universities. We evaluate whether articles associated with BRC deposits receive a "boost" in citations after deposit has occurred (after accounting for

an article-specific fixed effect, and controlling for vintage and year effects). In so doing, we 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 results provides strong empirical support for both the selection effect and the marginal impact of BRCs. Even in the period prior to their accessibility 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. Once the materials associated with these articles are made accessible through a BRC, the articles 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 is not merely persistent, but grows over time. This finding is consistent with the role of BRCs in helping to preserve the accessibility of knowledge for future research generations.

Finally, we conduct a "rate-of-return" calculation that compares the benefits (in terms of cost per citation) in making research accessible through a BRC versus simply funding additional research studies. Overall, we estimate that BRC deposit is associated with a three-fold efficiency gain in terms of inducing citations. While we are cautious in our interpretations are note that our analysis is subject to the caveats associated with any research premised on citation analysis, our results seem to provide evidence that the types of historical institutions emphasized by Mokyr (2002) seem to matter for cumulative progress in modern life sciences research.

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 (Heller

and Eisenberg, 1998; David, 2001; Mokyr, 2002). The diffusion of knowledge over time depends on institutions that facilitate low-cost knowledge transfer across 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 institutions on the extent of knowledge spillovers. This research often employs citations to academic papers or approved patents to estimate the influence of prior knowledge on current advances. Perhaps no research-enhancing institution has been more intensively studied than universities. 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

[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.

national laboratories (Jaffe and Lerner, 2001), venture capital (Kortum and Lerner, 2000), and patent pools (Lerner and Tirole, 2003), among others.[4]

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 has confounded 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) because the research reflected in the patent is more fundamental or because the norms of disclosure and 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[5]

III.A. What are Biological Resource Centers?

BRCs are institutions that collect, certify, and distribute biological organisms for use in life science research and in the development of commercial products in the pharmaceutical, agricultural and biotechnology industries. As a key element of the life sciences research infrastructure, BRCs maintain a large and varied collection of biological materials, including cell lines, micro-organisms, recombinant DNA material, biological media and reagents, and the information technology tools that allow researchers to access biological materials. Over the past quarter century, they have come to play an increasingly important role in scientific and commercial research. For example, since the 1980s, select BRCs have been critical to the

[4] 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).

[5] Stern (2004) provides a far 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.

extension of intellectual property rights, by serving as International Patent Depositories for all patented living organisms. By making decisions about which materials to store and by affecting the cost of access to research materials, BRCs play a potentially important role as a research-enhancing institution facilitating the diffusion of useful knowledge.

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

Several alternative institutional arrangements for collecting, certifying, and circulating biological materials exist: peer-to-peer networks, private culture collections, and for-profit culture firms. 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. Private collections, such as those within individual companies or universities, are less idiosyncratic but are dispersed and offer only minimal certification and quality assurances. While for-profit culture distribution firms may offer high-quality products, their incentives to maintain large collections over the long-term are limited. In contrast, BRCs are associated with distinctive institutional features to enhance the diffusion of scientific knowledge, including preservation, certification, independent access, and scale and scope economies. As discussed in the following sub-sections, each of these features is consistent with the attributes required for cumulativeness identified by Mokyr (2002).

III.B. 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 is illustrated dramatically through the case of PCR, perhaps the single most important research tool in biotechnology. 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). Though the costs to develop materials that would enable Mullis' technique might have taken years to develop, Cetus researchers contacted the ATCC. Since the mid-1960s, ATCC had maintained an *extremophile* collection, including micro-organisms such as *Thermus aquaticus* which had been initially discovered in the hot springs of Yellowstone National Park. Though no practical benefit was foreseen at the time the collection was established, the materials maintained in that collection were ideal for realizing Mullis' vision. Specifically, the (rather obscure) fact that the DNA polymerase associated with this organism maintained its enzymatic properties even during rapid heating and cooling allowed Cetus to 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. 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). Overall, where BRCs explicitly focus on preservation, there are few incentives within for-profit entities or the peer-to-peer network to maintain to 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.C. Certification in Biological Resource Centers

BRCs also serve to *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.[6]

Though seemingly straightforward, the certification function is critical to effective life sciences research. Consider the so-called "HeLa scandals" (Gold, 1986; Masters, 2002). Before researchers grasped the importance of biomaterials fidelity (and before verification techniques were widely understood), research materials were customarily exchanged between research labs through peer-based networks. Though most researchers assumed that biomaterials exchanges simply offered a low-cost procedure for building on discoveries made by others, this system

[6] The ATCC and DSMZ, for example, regularly issue notices identifying materials errors and misclassifications.

resulted in what many consider to be the most far-reaching methodological failure in modern life sciences research. In particular, without specific methods for certifying the authenticity of biomaterials, a surprisingly large share of the materials exchanged between labs became contaminated with other cell lines or otherwise misidentified. In other words, a scientist using materials from another laboratory might believe that an experiment was being conducted on cells from the healthy embryonic lung cells of a white male but was, in fact, unwittingly using cancerous cervix cells from a 31-year old black woman. In a series of revelations that came to be known as the "HeLa scandals" (named after Henrietta Lacks, a woman whose a exceptionally robust cervical cancer cell line was "responsible" for a large number of contaminations), certification errors cast doubt over thousands of individual research findings, including the research of Nobel Prize-winning scientists and other researchers around the world.[7]

The consequences of misidentification are far-reaching. Not only does misidentification cast a cloud over the findings of current researchers (with career implications for those whose results are under suspicion), but confusion and uncertainty places a longer-term cost on progress. Effective certification, as is made feasible by the activities of BRCs, allows researchers to build on the insights of prior research, avoid needless and costly duplication, and so 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.

III.D. 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

[7] 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.

encouraging *independent and open access* to the results of prior scientific research. In non-BRC networks, access to source materials is dependent on the "good will" of researchers who maintain active cell cultures within their laboratory; such goodwill is difficult to maintain when researchers are simultaneously competing with each other to establish new research findings or when 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 over the "ownership" of the results of the Human Genome Project are but the most visible in the ongoing war over access to biological materials and data.

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, damaging Gallo's reputation and delaying 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.E. Scale and Scope Economies

Finally, as "living libraries" that continuously collect material developed by the scientific community, BRCs are able to achieve substantial scale and scope economies. Relative to other organizational forms that preserve life science materials, BRCs maintain larger, more varied, and more balanced collections. As a result, BRCs are more likely to undertake the investments that are necessary to increase the quality and reduce the cost of accessing biological materials. For example, 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, BRC lower the costs to access 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. We therefore exploit the timing of transfer for some collections that had been maintained in a private university laboratory that get exogenously 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 (exogenously timed) BRC deposits to potentially citable scientific research articles, we implement a differences-in-differences estimator of the marginal impact of BRC deposit.[8] Specifically, we construct a dataset composed of scientific publications linked to

[8] 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 find no statistically significant evidence for such a trend.

(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.[9] 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.

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 fixed effects.[10]

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:

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

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

[9] In particular, the two control groups are (a) the research article immediately preceding the BRC-linked article in the journal in which it was published and (b) the research article ranked "most related" by the online search tool PUBMED in the (annual) volume of the journal in which the BRC-linked article is published.

[10] Several subtle issues, including an 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, Griliches, and Hall (1984) and the fixed effects estimator suggested by Allison and Waterman (2001). All of our qualitative findings are unchanged across these different procedures; building on the simulation results reported by Allison and Waterman (2001), we report the results associated with the fixed effects procedure.

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.[11] 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. To be conservative, we therefore examine (and base our core policy findings on) a series of estimates including article-specific fixed effects (γ_i), as in the following specification:

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

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:

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

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} is increasing 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 interact POST-DEPOSIT with technological and institutional factors associated with the scientific research article and deposit itself. For example, we exploit the variation in data that results from the transfer of three distinct biological materials collections to be shifted from the peer-to-peer network to the ATCC, the leading BRC in the United States. Consistent with our prior discussion, we can estimate the separate impact of each of these deposit events on subsequent citation to linked research articles. 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.

[11] 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.

V. Data

V.A. Data Construction and Sources

To implement the differences-in-differences strategy described above, we address four main challenges: (a) linking BRC deposits to research publications, (b) identifying a sample of publications which 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 largest BRC in the United States, 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.[12]

To overcome the second challenge, we take advantage of exogenous shocks that lead to the bulk transfer of materials into ATCC from other collections. In particular, our dataset is based on a series of three "special collections" maintained by the ATCC. These materials transfers occurred when scientists who maintained collections within the peer-to-peer network retired, moved or faced an institutional funding limitation 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 funding considerations and was accessioned beginning in 1982. Seventy-seven articles associated with 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 funding considerations led to its being transferred into ATCC beginning in 1981. Forty-four articles from 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

[12] 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.

left his position as Head of Tumor Cell Biology Section at the National Cancer Institutes, along with his collaborator, Dr. John Minna, to take a position at UT Southwestern. The materials in the Gazdar collection were accessioned beginning in 199, linked to six research articles.

The shift of "special collections" from the peer-to-peer network into a BRC constitutes a (testably) exogenous event that enables us to test the impact of alternative institutional arrangements on the diffusion of knowledge. 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. The three collections we examine were chosen to avoid potential endogeneity. Specifically, we considered several alternative collections; however, in these other cases, BRC deposit occurred as the result of the development of scientific or technological interest in a particular class of materials that had been maintained in the peer-to-peer network (and so using these collections would lead to spurious correlation between deposit and a citation "boost"); in contrast, the "trigger event" leading to deposit for the three collections included in here is unrelated to the intrinsic value of the knowledge associated with the materials.[13]

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).[14] 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.[15] By matching control articles to treatment articles in this way, we attempt to minimize heterogeneity associated

[13] The history of ATCC collections is drawn from discussions with Dr. Raymond Cypess, President and CEO of the ATCC, and Dr. Robert Hay, Director of the Department of Cell Biology at ATCC, and other ATCC staff members.
[14] 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.

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. By 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).[16] 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.[17]

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, 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

---

[15] When the ATCC-associated article is the lead article, we use the second article in that issue as the control.
[16] 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.

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, year of, and 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 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.

V.B. Summary Statistics

Table 1 provides variable names and definitions and Table 2 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 A. 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 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 B). By the end of 2001, the average article in our sample has received more than 81 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

[17] 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.

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. For each of the materials in the special collections, we also track the current PRICE; this averages approximately $230 per order.

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.

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

Table 3 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 C-1 and C-2 portray the disparity between these groups over time, comparing average citations by article age for each control group. Figure C-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 C-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 patterns 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 relying (or excluding) 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 pairs* (i.e., comparing articles eventually deposited in BRCs with those that are not) and *across deposit-status within article* (i.e., whether a particular article has yet been deposited), we can precisely identify the marginal impact of BRC deposit on the rate of knowledge diffusion. In particular, after controlling for other factors, a positive and significant effect on BRC ARTICLE implies a selection effect, while a positive and significant effect on BRC ARTICLE, POST-DEPOSIT is the estimated marginal impact of BRCs. 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 4, 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. Equation (4-1) and (4-2) presents 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 control articles), and receive an additional 64.4% increase in their citation rate after BRC deposit. BRC-linked articles also experience a citation boost during the WINDOW PERIOD, although at 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 receives an additional "boost" in the years following BRC deposit.

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 fixed effects negative binomial specification in the remainder of the analysis. We report 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).[18] The first of these specifications (4-3) is 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 150% *after* BRC deposit).

Additional evidence is presented in (4-4) and (4-5), where we control for observable characteristics of articles and collections. For example, as presented in (4-4), articles associated with the HTB, TIB, and Gazdar collections are 68%, 123%, and 79% more likely to be cited than controls (and the impact of BRC ARTICLE, POST-DEPOSIT remains large and positive). As

[18] As a descriptive exercise, we first implement a negative binomial specification with article-specific fixed effects, article age effects, and calendar year effects. (all of which are highly significant). Figures D-1 and D-2 portray the conditional article age and article fixed effects, respectively. Even after accounting for the other effects, Figure D-1 suggests that article citation rates first rise and peak in three to four years after of publication and then decline over time. Figure D-2 confirms that there exists substantial variation across articles in the "baseline" citation rate.

well, in (4-5), we investigate the impact of other factors that might affect the citation rate; with the exception of CUMULATIVE CITATIONS (i.e., citations between the publication year and year *t-1)* none of the other article characteristics – UNIVERSITY LEAD AUTHOR, GOVERNMENT LEAD AUTHOR, FOREIGN LEAD AUTHOR, and PAGES – impacts citations in a statistically significant manner. It is worthwhile to note that the inclusion of CUMULATIVE CITATIONS does not obviate the impact of BRC-association or the impact of deposit. This suggests that the findings hold equally for highly cited and less highly cited articles. Overall, the findings in Table 4 provides evidence that accords with the presence of both a selection effect and the marginal impact of BRCs on the diffusion of scientific knowledge.

VI.B. Fixed-Effects Specifications

While Table 4 separately isolates the selection and marginal impact of BRCs, we have so far ignored the substantial variability among articles, even within article families. While specifications that include article fixed effects do not identify the average selection effect, the use of article fixed effects (as in Table 5) does 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). As demonstrated in (5-1), the average article in estimated to experience a 124% citation boost after BRC accession (and a 43% boost in citations during the WINDOW PERIOD), even after controlling for all article, age and year-specific effects. Moreover, this finding is robust to the inclusion of interactions with observable article characteristics. Similar to (4-5), (5-2) finds no evidence for the influence of foreign authorship, non-university authorship, or top-university on the citation boost. However, the impact of deposit does seem to vary by collection. Articles associated with the HTB and TIB collections experience post deposit citations boosts of 185% and 121%, respectively. In contrast, there is no significant impact for articles in the Gazdar collection (though this collection was only accessioned in 1994 and contains only a small number of articles).

Of course, 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 6, 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. As well, BRC-ARTICLE, POST-DEPOSIT remains positive, significant, and of an important economic magnitude (38.6%). While BRC-affiliated articles do experience an additional upward citation trend, these articles experience a post-deposit citation boost even when incorporating this trend into the analysis.

Simply allowing a BRC-linked time trend is inadequate, however, for two reasons. First, if the time trend is particularly pronounced in the years just prior to initial 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 support for 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 E. 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 E presents a specification similar to Equation (3), but including a 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 E plots each coefficient (in terms of the incidence-rate ratio minus 1), excluding the years associated with the accession window (all effects are computed relative to the window period). Two findings stand out. First, there is no significant pre-deposit trend (e.g., the coefficient associated with the dummy five years prior to deposit indicates a higher citation rate than any of the coefficients in the years just prior to deposit). Second, there is a substantial – 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 continuing 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.

VI.C. The Impact of Deposit over the Sample Period

Figure F further considers how the impact of BRC deposit has changed over time. This graph reports the coefficients from a regression similar to (5-1) that includes dummy variables for the interaction between each calendar year since 1984 and POST-ATCC-DEPOSIT. The value of BRC deposit is statistically significant in all years, and the impact of BRC association has steadily increased over time, particularly in the 1990s. This is consistent with the idea that BRCs have become increasingly effective over time in contributing to the diffusion of knowledge associated with deposited materials.

VI.D. Robustness to Alternative Control Groups and Specifications

Table 7 explores the robustness of the results, examining the inclusion or exclusion of alternative control groups and the exclusion of the data associated with the window period. Following the basic specification implemented in (5-1), (7-1) and (7-2) 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 "nearest neighbor" control. In (7-3), we simply drop the data associated with the window period. The results are essentially unchanged.

VI.E. Assessing the Cost-Effectiveness of BRCs

Our final exercise is to undertake a "back-of-the-envelope" cost-effectiveness analysis. A comprehensive cost-benefit analysis is beyond the scope of our analysis, since it is not possible to fully capture the degree to which BRC materials actually change the productivity of research by users. However, we can make a simple comparison between the efficacy of funding the marginal investment in accessioning materials to be included in a BRC versus using that same budget to simply fund additional research. In both cases, expenditures will be associated with an increase in the knowledge available to future researchers, which will be reflected in citations to specific research articles. Our cost-effectiveness analysis compares the predicted impact on future citations of providing additional funds for research studies versus providing those funds for investments to ensure that today's discoveries are accessible to follow-on researchers. To undertake this calculation, we must compute a cost per citation associated with funding research studies, the cost for BRC accession, the number of citations induced by BRC

accession. Together, these figures will allow us to compare the cost per citation of funding additional research versus ensuring the accessibility of research.

The first step is obtaining a baseline citation cost, i.e., the "cost per citation" paid by public funding agencies (such as NIH) when allocating resources that result in published scientific articles. We draw on the estimates of Adams and Griliches (1998). Using data from the 1980s, Adams and Griliches estimate the relationship between expenditures and academic research output (papers and citations) for individual academic departments at top universities across the United States, including biology departments. Using these measures (and converting expenditures into 1987 current dollars), they estimate the cost per citation to be $2400 for expenditures at a top-ten biology department and at $4200 for citations at non-elite public universities. Using the BEA R&D price deflator to restate this figure in current dollars, the *lowest* Adams and Griliches estimates of cost per citation is $2887. Being conservative (in terms of estimating the effectiveness of BRC expenditures), we choose the lowest estimated cost per citation among these figures, and so set the Baseline Citation Cost at $2887 for the life sciences.

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

Third, we calculate the incremental number of citations expected to result from deposit and accession into a national BRC. To calibrate our expectation for the impact of deposit on citation, we use the estimated POST-DEPOSIT impact of 56.5% obtained in (6-2). This estimate is more conservative than nearly all others in our analysis, and, to be conservative, we also omit the impact of the post-deposit trend included in that specification.

We use the 56.5% figure to compute four alternative estimates of the cost per citation associated with BRC deposit. The first two of these computations build on the data provided by Adams and Griliches (1998). In their work, the average biology publication received 24.6 citations during the first five years of publication if authors were located at a top ten university and 14.3 citation if authors were located at universities below the top ten (in biology). Were a

material associated with a publication from a Top Ten university accessioned, we would expect that the associated article would receive a citation boost equal to 13.90; were the accessioned materials drawn from a random university, we would expect the citation impact to be 11.7, based on the citation rates of articles published by authors outside the top ten. The second pair of estimates are based on the incremental citations that would be associated with articles in our data sample. If we include the entire sample (including the control articles), the incremental citations over the first five years after publication is estimated to be 23.90. If we condition on those articles which are ultimately linked to a BRC, the estimated citation boost increases to 36.83.

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

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 a given level of research from one generation as “seed corn” for future researchers.

Our principal contribution has been to provide evidence about the role of BRCs as a quite specific (and somewhat "invisible") institution within the life sciences that impact the cumulativeness of research in that field. We separated out a selection effect (which turned out to be quite important) from the marginal impact of the institution in enhancing knowledge diffusion. 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 are able to find evidence in favor of the impact of BRCs, our results have not, by and large, separated out the differences between the mechanisms associated with research-enhancing institutions. Our results on the growing importance of the BRC marginal effect in the time elapsed since deposit is consistent with the role of a preservation function. However, we do not differentiate between the role of certification versus the role of independent access for follow-on researchers. 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 disentangle these two different functions of BRCs. Second, whereas our econometric exercise exploits specific instances associated with "exogenous" deposit, 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. In Mukherjee and Stern (2004), we provide a simple overlapping generations model that identifies the key factors underlying this endogenous choice, and derive the conditions under which investments in institutions that support disclosure and cumulativeness provide a social welfare gain. 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 influence 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 surrounding 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 generations.

Finally, our empirical approach highlights an important but often over-looked 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 Liebskind, 1998; David, 2001; Murray, 2002). A fruitful area for future research, then, is to identify a setting similar to that considered in this paper that would allow us to infer the marginal impact of university policies (or intellectual property policies, for that matter) on the commercialization of knowledge. Such a study will allow an economic analysis of the conditions supporting different types of knowledge production at a point in time and the conditions supporting knowledge accumulation over time.

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. 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. Liebskind (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.

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.

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.

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

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

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

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.

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,

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

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.

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

TABLE 1
VARIABLES & DEFINITIONS

VARIABLE	DEFINITION	SOURCE
CITATION CHARACTERISTICS		
FORWARD CITATIONS$_{jt}$	# of Forward Citations to Article *j* in Year *t*	Science Citation Index (SCI)
CUMULATIVE CITATIONS$_{jt}$	# of FORWARD CITATIONS from publication date to YEAR$_{t-1}$	SCI
YEAR	Year	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

TABLE 2
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	1999.79	7.21	1970	2001
AGE	6475	11.27	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

* *DEPOSIT YEAR & PRICE are only meaningful for articles associated with BRC deposits.*

^ Institutional affiliations are not available for some of the publications.

TABLE 3
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 4
BASELINE SPECIFICATION*

	OLS Dep Var = ln(FORWARD CITATIONS)		**NEGATIVE BINOMIAL *(Coefficients reported as incidence-rate ratios)* Dep Var = FORWARD CITATIONS**		
	(4-1) Base Model: BRC Effect with Age FEs only	**(4-2)** Base Model: Including Article Family & Year FEs	**(4-3)** Baseline Count Model	**(4-4)** with Collection Effects	**(4-5)** *(4-3)* with Article Characteristics
ARTICLE CHARACTERISTICS					
BRC-ARTICLE	**0.465 (0.143)**	**0.513 (0.122)**	**2.013 (0.312)**		**1.864 (0.298)**
BRC-ARTICLE, WINDOW PERIOD	**0.351 (0.122)**	**0.355 (0.105)**	**1.486 (0.181)**	**1.450 (0.183)**	**1.274 (0.179)**
BRC-ARTICLE, POST-DEPOSIT	**0.644 (0.161)**	**0.574 (0.132)**	**2.504 (0.430)**	**2.451 (0.436)**	**1.863 (0.320)**
HTB-ARTICLE				**1.683 (0.407)**	
TIB-ARTICLE				**2.231 (0.450)**	
GAZDAR-ARTICLE				**1.791 (0.394)**	
CUMULATIVE CITATIONS					**1.004 (0.001)**
UNIVERSITY LEAD AUTHOR					0.953 (0.186)
GOVERNMENT LEAD AUTHOR					1.037 (0.251)
NON-US LEAD AUTHOR					1.061 (0.227)
NUMBER OF AUTHORS					1.011 (0.018)
PAGES					0.996 (0.019)
CONTROL VARIABLES					
Parametric Restrictions					
Article Family FEs = 0 *(#restrictions = 108)*		**F-stat 7.8*10^11 p-value 0.00**	**χ^2 5.0*10^11 p-value 0.00**	**χ^2 5.4*10^10 p-value 0.00**	**χ^2 2.7*10^7 p-value 0.00**
Age FEs = 0 *(#restrictions = 30)*	**F-stat 16.84 p-value 0.00**	**F-stat 14.77 p-value 0.00**	**χ^2 224.56 p-value 0.00**	**χ^2 221.18 p-value 0.00**	**χ^2 267.29 p-value 0.00**
Year FEs = 0^ *(#restrictions = 23)*		**F-stat 2.81 p-value 0.00**	**χ^2 53.23 p-value 0.00**	**χ^2 53.18 p-value 0.00**	**χ^2 77.41 p-value 0.00**
REGRESSION STATISTICS					
R-squared	0.26	0.51			
Log-likelihood			-14350.77	-14344.37	-10758.49
# of Observations	6475	6475	6475	6475	4715

* 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 5
ARTICLE FIXED EFFECTS SPECIFICATION*

	NEGATIVE BINOMIAL *(Coefficients reported as incidence-rate ratios)* **Dep Var = FORWARD CITATIONS**		
	(5-1) Core Article Fixed Effects Model	**(5-2)** *(5-1)* with Post-Deposit *Article Characteristics	**(5-3)** Results by Collection
ARTICLE CHARACTERISTICS			
BRC-ARTICLE, WINDOW PERIOD	**1.432 (0.181)**	**1.448 (0.181)**	**1.450 (0.185)**
BRC-ARTICLE, POST-DEPOSIT	**2.242 (0.373)**	**2.053 (0.638)**	
HTB-ARTICLE, POST-DEPOSIT			**2.852 (0.686)**
TIB-ARTICLE, POST-DEPOSIT			**2.212 (0.409)**
GAZDAR-ARTICLE, POST-DEPOSIT			1.238 (0.313)
BRC-ARTICLE, POST DEPOSIT * NON-US AUTHOR		1.655 (0.522)	
BRC-ARTICLE, POST DEPOSIT * NON-UNIVERSITY AUTHOR		1.098 (0.318)	
BRC-ARTICLE, POST DEPOSIT * TOP-UNIVERSITY AUTHOR		1.412 (0.491)	
CONTROL VARIABLES			
Parametric Restrictions			
Article FEs = 0 *(#restrictions = 287)*	**F-stat $2.4*10^{12}$ p-value 0.00**	**F-stat $1.3*10^{7}$ p-value 0.00**	**F-stat $5.2*10^{11}$ p-value 0.00**
Age FEs = 0 *(#restrictions = 30)*	**F-stat 303.33 p-value 0.00**	**χ^2 254.43 p-value 0.00**	**χ^2 302.44 p-value 0.00**
Year FEs = 0^ *(#restrictions = 23)*	**F-stat 50.44 p-value 0.00**	**χ^2 42.18 p-value 0.00**	**χ^2 48.51 p-value 0.00**
Regression Statistics			
Log-likelihood	-12750.84	-10349.43	-12741.85
# of Observations	6475	4892	6475

* 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 6
BRC-DEPOSIT ARTICLE TIME TRENDS*

	NEGATIVE BINOMIAL REGRESSIONS *(Coefficients reported as incidence-rate ratios)* **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	**1.201 (0.159)**	**1.537 (0.245)**
BRC-ARTICLE, POST-DEPOSIT	**1.386 (0.228)**	**1.565 (0.321)**
BRC-ARTICLE * TIME TREND	**1.045 (0.013)**	
BRC-ARTICLE * PRE-DEPOSIT TREND		1.045 (0.245)
BRC-ARTICLE * POST-DEPOSIT TREND		**1.037 (0.122)**
CONTROL VARIABLES		
Article FEs = 0 *(#restrictions = 287)*	**χ^2 4.9*10^11 p-value 0.00**	**χ^2 5.0*10^11 p-value 0.00**
Age FEs = 0 *(#restrictions = 30)*	**χ^2 300.44 p-value 0.00**	**χ^2 290.94 p-value 0.00**
Year FEs = 0^ *(#restrictions = 23)*	**χ^2 53.70 p-value 0.00**	**χ^2 54.17 p-value 0.00**
Regression Statistics		
Log-likelihood	-12719.30	-12727.65
P-value of Chi	0.00	0.00
# of Observations	6475	6475

* 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 7
ROBUSTNESS TO ALTERNATIVE CONTROL GROUPS AND SPECIFICATIONS*

	NEGATIVE BINOMIAL REGRESSIONS *(Coefficients reported as incidence-rate ratios)* **Dep Var = FORWARD CITATIONS**		
	(7-1) Only "Nearest Neighbor" Controls	**(7-2)** Only "Most-Related Article" Controls	**(7-3)** Omitting Window Period Observations
ARTICLE CHARACTERISTICS			
BRC-ARTICLE, WINDOW PERIOD	**1.381** **(0.177)**	**1.304** **(0.173)**	
BRC-ARTICLE, POST-DEPOSIT	**2.222** **(0.389)**	**1.612** **(0.313)**	**2.188** **(0.377)**
CONTROL VARIABLES			
Article FEs = 0 *(#restrictions = 287)*	**χ^2 2.2*10^7 p-value 0.00**	**χ^2 5.1*10^11 p-value 0.00**	**χ^2 1.6*10^7 p-value 0.00**
Age FEs = 0 *(#restrictions = 30)*	**χ^2 386.01 p-value 0.00**	**χ^2 216.26 p-value 0.00**	**χ^2 259.16 p-value 0.00**
Year FEs = 0^ *(#restrictions = 23)*	**χ^2 58.32 p-value 0.00**	**χ^2 59.71 p-value 0.00**	**χ^2 42.42 p-value 0.00**
Regression Statistics			
Log-likelihood	-9788.17	-9280.52	-11801.43
P-value of Chi	0.00	0.00	0.00
# of Observations	4833	4060	6165

* 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 8
BRC DEPOSIT COST-EFFECTIVENESS ANALYSIS

Calculation	Baseline Cost Per Citation*	BRC Accession Cost	BRC Citation Boost	Cost per Citation for BRC-affiliated Articles	BRC Cost-Effectiveness Index¥
BRC-Deposited Articles Citation Boost	$2,887	$10,000	36.83	$271.52	10.63
Sample Article Citation Boost	$2,887	$10,000	23.90	$418.41	6.90
Top Ten University Citation Boost^	$2,887	$10,000	13.90	$719.48	4.01
Random University Citation Boost^	$2,887	$10,000	8.08	$1237.70	2.81

***** Based on Adams-Griliches (1998) estimate of cost per citation.

^ Based on Adams-Griliches (1998) estimate of citations received by articles authored by member of Top Ten Biology departments and other university Biology departments.

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

FIGURE A
NUMBER OF PUBLICATIONS BY YEAR

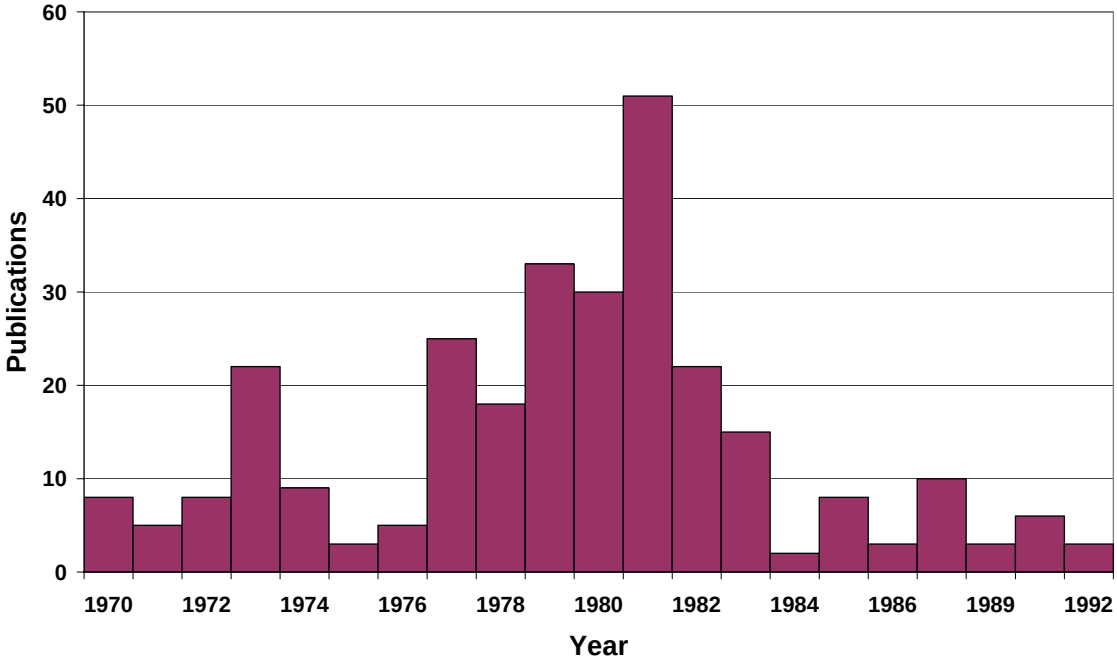


FIGURE B
DISTRIBUTION OF CITATIONS

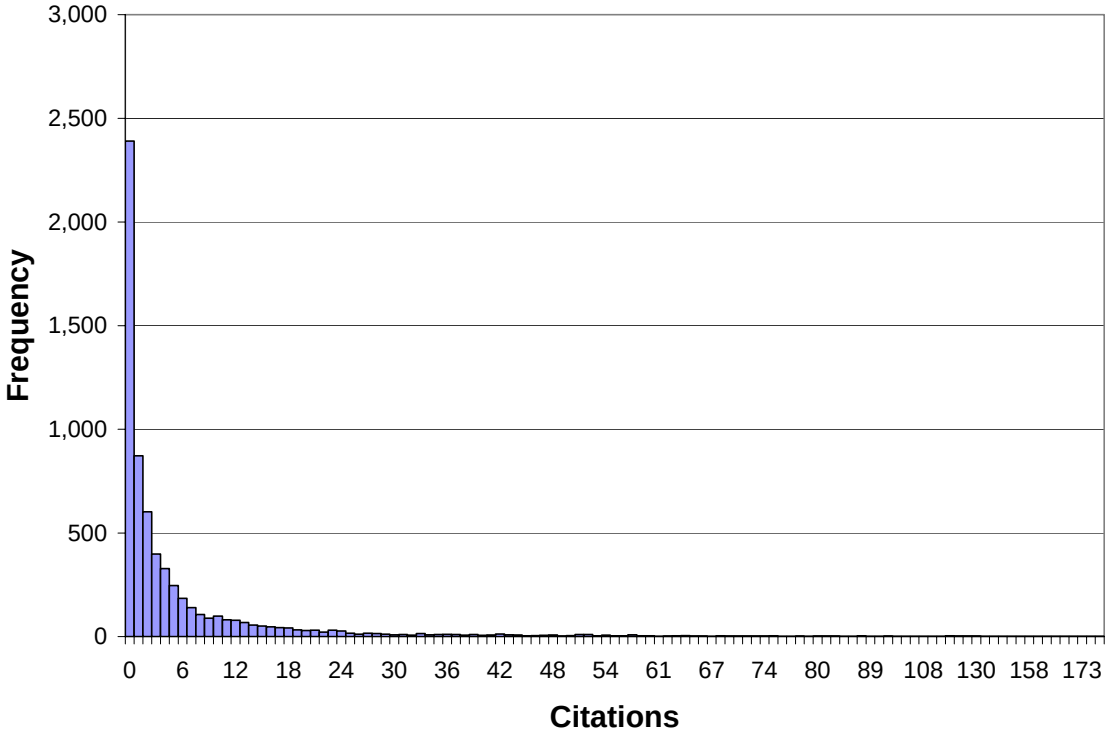


FIGURE C-1
AVERAGE ANNUAL CITATIONS BY AGE, BRC VS. CONTROL ARTICLES

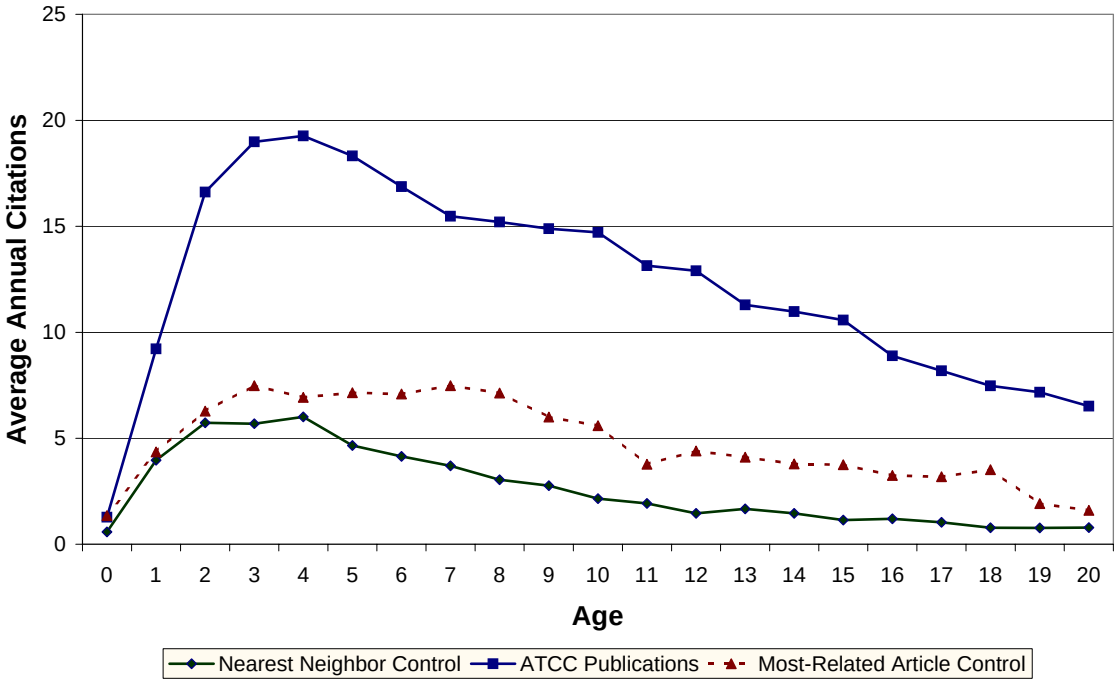


FIGURE C-2
PERCENT DIFFERENCE IN ANNUAL AVERAGE CITATIONS BY AGE, BRC VS. CONTROL ARTICLES

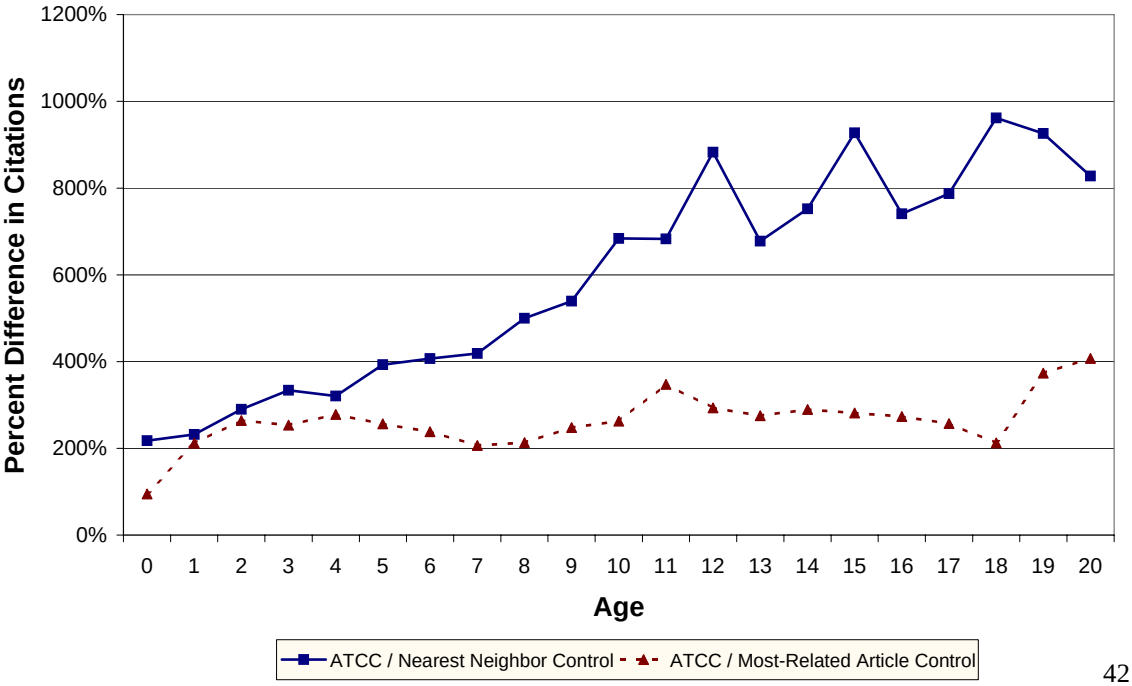
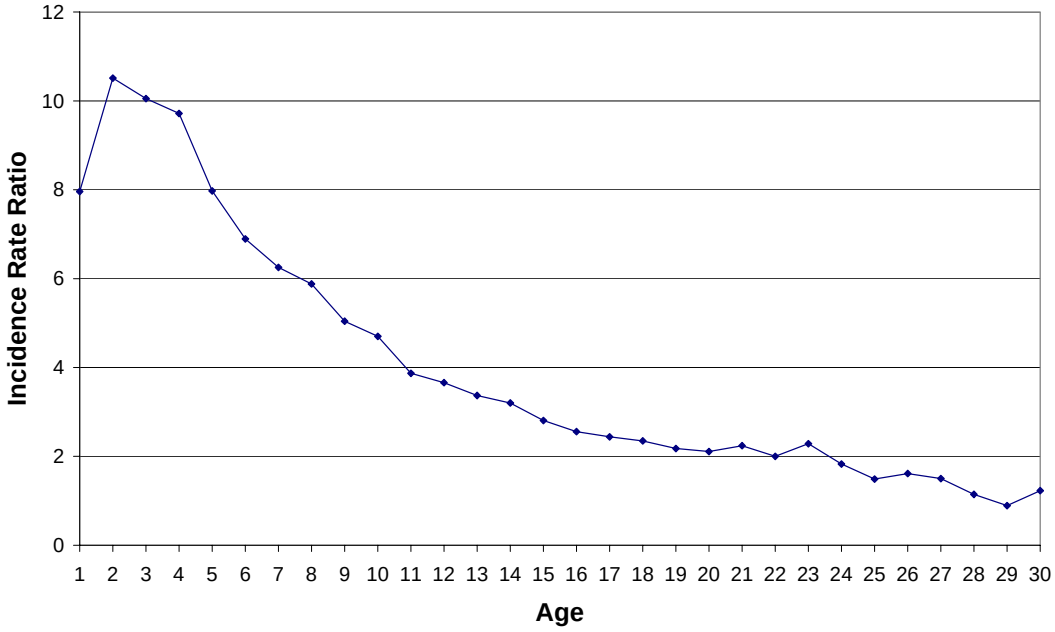


FIGURE D-1
CONDITIONAL AGE EFFECTS



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

FIGURE D-2
CONDITIONAL PAPER EFFECTS

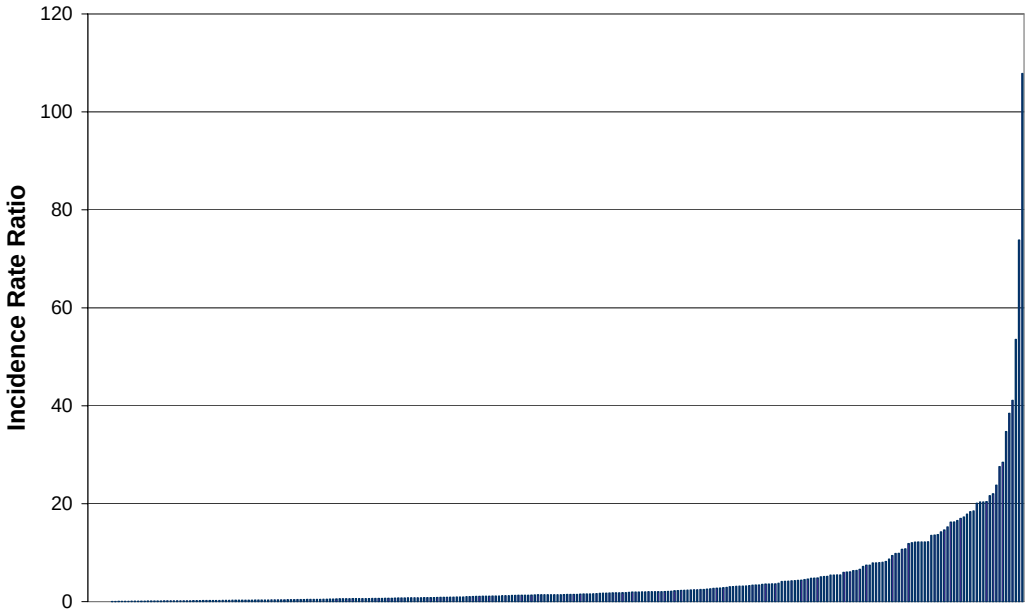


FIGURE E
PRE- AND POST-DEPOSIT EFFECTS ON FORWARD CITATIONS

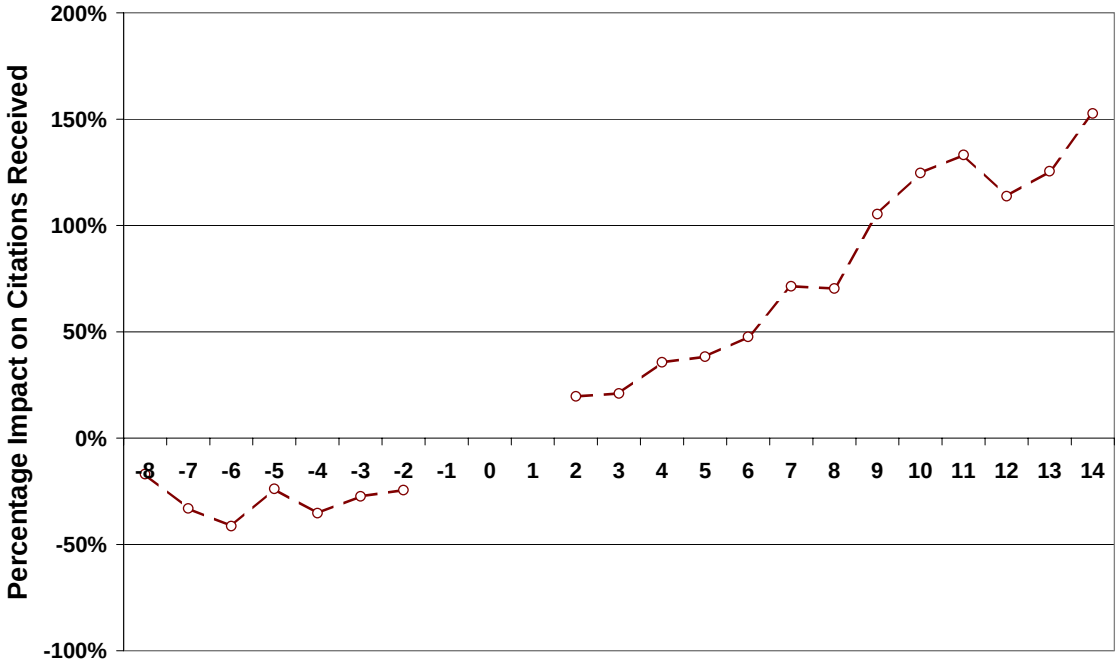


FIGURE F
IMPACT OF BRC DEPOSIT ON FORWARD CITATIONS, MARGINAL EFFECTS BY YEAR