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

Citation: Furman, Jeffrey L, and Scott Stern. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." American Economic Review 101.5 (2011): 1933–1963.

As Published: http://dx.doi.org/ 10.1257/aer.101.5.1933

Publisher: American Economic Association

Persistent URL: http://hdl.handle.net/1721.1/75830

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

American Economic Review 101 (August 2011): 1933–1963
http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.1933

Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research[†]

By Jeffrey L. Furman and Scott Stern*

While cumulative knowledge production is central to growth, little empirical research investigates how institutions shape whether existing knowledge can be exploited to create new knowledge. This paper assesses the impact of a specific institution, a biological resource center, whose objective is to certify and disseminate knowledge. We disentangle the marginal impact of this institution on cumulative research from the impact of selection, in which the most important discoveries are endogenously linked to research-enhancing institutions. Exploiting exogenous shifts of biomaterials across institutional settings and employing a difference-in-differences approach, we find that effective institutions amplify the cumulative impact of individual scientific discoveries. (*JEL* D02, D83, I23, O30)

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

— Isaac Newton, 1676

At least since the development of scientific societies and related research institutions in the seventeenth century, the centrality of cumulative knowledge in scientific and technical advances has been recognized.[1] However, from the perspective of economic theory, knowledge accretion has been incorporated only recently, through models of endogenous economic growth (Paul M. Romer 1990; Gene M. Grossman

* Furman: Boston University School of Management, 595 Commonwealth Ave. #653a, Boston, MA 02215, and NBER (e-mail: furman@bu.edu); Stern: MIT Sloan School of Management, E62-476, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: sstern@mit.edu). 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, Iain Cockburn, Amy Finkelstein, Joshua Gans, Shane Greenstein, Bronwyn Hall, Ben Jones, Michael Kremer, Julia Lane, Robert Litan, Megan MacGarvie, Fiona Murray, Paul Romer, and participants in numerous seminars, for comments and suggestions. Brigitte Adam, Lisa Bassett, Peter Becker, Ann-Marie Crain, Jason Corradini, Mercedes Delgado, Lorraine DeLeon, Chijoke Emineke, Devin Fensterheim, Pegi Fogertey, Jeremy Gagne, Anna Harrington, Martha Kam, Hartmut Klipsch, Julia Lo, Kinga Piekos, Brianna Pesci, Courtney Mason, Tracy Myers, Jochem Rogowicz, Mathieu Trepanier, Michael Vitulli, and Naomi Wohl provided excellent research assistance. We also thank Jim Vaught of the National Cancer Institute and Lisa Miranda for their insights on biospecimen collections. All errors are our own. Financial support for this research was provided by the National Science Foundation Science of Science Policy Program (grant 0738394), the Brookings Institution, and the Boston University School of Management Junior Faculty Research Fund.

† To view additional materials, visit the article page at http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.1933.

[1] Isaac Newton famously acknowledged the importance of cumulative research in a 1676 letter to rival Robert Hooke: "What Descartes did was a good step. You have added much several ways, and 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" (quoted in Stephen Inwood, 2003, p. 216).

and Elhanan Helpman 1991; Philippe Aghion and Peter Howitt 1992; Charles I. Jones 1995) and step-by-step technical progress within industries (Suzanne Scotchmer 1991; Nancy Gallini and Scotchmer 2002; Aghion, Mathias Dewatripont, and Jeremy C. Stein 2008). In order to serve as a foundation for long-term growth, scientific research and technological progress must exert a positive intertemporal spillover; to avoid diminishing returns to research investments, research itself must "stand on the shoulders" of prior knowledge (Jones 1995).

Though extremely insightful in deriving the implications of knowledge accumulation for related economic variables (such as the equilibrium growth rate or the incentives for innovation), these models do not articulate the conditions that facilitate knowledge accumulation. As Joel Mokyr (2002) 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 be aware of extant knowledge and pay the associated costs of access. Further, since any one researcher captures a small share of the benefit from the process of certifying knowledge and making it accessible, there may be a significant gap between the private and social returns associated with investments that contribute to the diffusion of scientific knowledge. Overall, the ability of a society to stand on the shoulders of giants depends not only on generating knowledge, but also on the quality of mechanisms for storing, certifying, and accessing that knowledge.

Institutions and public policy are often suggested as central to the process of knowledge accumulation.[2] Social scientists face a considerable challenge, however, in assessing the extent to which any one institution influences the creation, maintenance, and extension of the knowledge stock. It is empirically difficult to isolate the intrinsic impact of a particular piece of knowledge from the impact of the institutions in which it is embedded, although the two are conceptually distinct. While we are interested in the *marginal* impact of an institution, the incremental influence of that institution on knowledge accumulation (conditional on the nature and quality of knowledge associated with it), a *selection* effect may confound our analysis if knowledge of high intrinsic importance is endogenously embedded within "high-quality" institutions.

The main contribution of this paper is to provide direct statistical evidence of the marginal impact of a specific institution—biological resource centers (BRCs)—on knowledge accumulation. BRCs collect, certify, and distribute biological organisms, such as cell lines, microorganisms, and DNA material. The ability to exploit prior research in the life sciences depends on access to the cells, cultures, and specimens used in that research. Distinct among institutional arrangements for obtaining materials for research purposes, BRCs have the explicit objective of enhancing cumulative knowledge production through biomaterials preservation, certification, and circulation. Our analysis, therefore, evaluates whether the ability to access biomaterials through a BRC amplifies the impact of the scientific research that initially described those research materials.

[2] The role of institutions in scientific research is central to the sociology of science (Robert K. Merton 1973) and the "new" economics of science (Partha Dasgupta and Paul A. David 1994). The linkage between institutions and knowledge accumulation has long been emphasized in the economics of technical change (Vannevar Bush 1945; R. R. Nelson 1959; Nathan Rosenberg 1963, 1979; Nelson 1993; David 2001; Mokyr 2002).

Our approach extends citation analysis to investigate the impact of institutions on the dynamics of cumulative scientific discovery (Adam B. Jaffe, Manuel Trajtenberg, and Rebecca Henderson 1993; Zvi Griliches 1998). We exploit three aspects of our empirical setting to develop and implement a difference-in-differences estimate of the impact of BRCs on knowledge accretion. First, each piece of material deposited in a BRC is associated with a journal article that describes its initial characterization and application. Second, for specific types of BRC deposits, there is a significant lag between the initial article and the date of its deposit into a BRC; in certain cases, materials associated with "special collections" were transferred exogenously from smaller collections into a major BRC for reasons unrelated to the extent of their use. Third, detailed bibliometric data for the BRC-linked articles, a sample of control articles, and all of the articles citing these original research articles allow us to capture variation in the extent to which knowledge diffuses across different economic and institutional contexts.

Our empirical analysis focuses on whether articles associated with materials exogenously shifted into a BRC receive a boost in citations after their deposit into the BRC, controlling for article-specific fixed effects and fixed effects for article age and calendar year. Our setting allows us to evaluate both models that include a control sample and models that rely exclusively on variation in the timing and date of the "treatment" of the deposit of the biomaterial into the BRC. Both approaches provide evidence for the marginal impact of BRCs on subsequent knowledge; the post-deposit citation boost is estimated to be between 57 percent and 135 percent across different specifications. Empirical checks of our key identification assumptions reinforce our overall findings. We find that the marginal impact of BRC deposit is marginally higher for articles published in less prestigious journals and that the citation boost is concentrated in follow-on research articles involving more complex subject matter. Overall, the evidence suggests that, relative to alternative institutions, BRCs play a significant role in the accumulation of knowledge in the life sciences.

I. The Impact of Research-Enhancing Institutions on the Accumulation of Knowledge

The dynamic accumulation of knowledge has become a central issue to many areas of research. The diffusion of knowledge among researchers and across generations depends on institutions and policies that facilitate low-cost knowledge transfer. 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 [knowledge stock about natural phenomena and regularities]" (Mokyr 2002, p. 8). We describe economic institutions that promote the accumulation of knowledge through these mechanisms as *research-enhancing institutions*.

Over the past two decades, a significant body of research has investigated specific research-enhancing institutions, documenting the presence of (and recognizing the difficulties of estimating) knowledge spillovers (Griliches 1990).[3] This

[3] The "search for spillovers" includes studies of university policy (David C. Mowery et al. 2001), intellectual property (IP) policy (Mariko Sakakibara and Lee Branstetter 2001), R&D consortia (Douglas A. Irwin and Peter J. Klenow 1996; Branstetter and Sakakibara 2002), national laboratories (Jaffe and Josh Lerner 2001), venture capital (Samuel Kortum and Lerner 2000), patent pools (Lerner and Jean Tirole 2004), scientific research networks (Walter

research often employs citations to academic papers or granted patents to estimate the influence of prior knowledge on current advances. For example, Jaffe, Trajtenberg, and Henderson (1993) and Henderson, Jaffe, and Trajtenberg (1998) examine whether university patents receive citations at a higher rate and with greater geographical scope than "control" patents drawn from similar geographic and technological areas. 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), this prior research has not been able to disentangle whether these institutions facilitate knowledge accumulation *per se* or whether they are simply linked to knowledge that has a higher intrinsic impact. In the terminology of the program-evaluation literature, these prior studies conflate the marginal impact of research-enhancing institutions with the selection effect of knowledge into research-enhancing institutions. For example, university patents may be highly cited (relative to a control group of patents generated by private-sector laboratories) because the research reflected in the patent is more fundamental or because the norms of disclosure and openness associated with a university amplify the diffusion of university-generated knowledge.

In addition to having an impact on the extent to which knowledge diffuses, institutions can influence the *types* of projects, drawing on particular pieces of knowledge. For example, researchers who are pursuing more fundamental (or complex) breakthroughs and whose research is itself likely to receive a high level of scrutiny (e.g., by being published in a more prestigious journal) are more likely to draw upon knowledge that is embedded within research-enhancing institutions. The long-term impact of knowledge creation depends not only on its fundamental importance, but also on whether it is embedded in institutions that facilitate low-cost knowledge diffusion.

II. BRCs and Cumulative Research in the Life Sciences[4]

A central challenge in the biological sciences is the need to maintain the integrity of biomaterials and data while sharing these materials across researchers and over time. Problems associated with biomaterials fidelity have bedeviled the life sciences research community. For example, Walter Nelson-Rees and his collaborators documented that dozens of cell lines widely used in the 1970s had been contaminated by a particularly strong cell line known as *HeLa*, shedding doubt on decades of cancer research, including the work of Nobel laureates (Michael Gold 1986; Rebecca Skloot 2010).[5] Any uncertainty about biomaterials fidelity can result in considerable research delays, as scientists must undertake substantial efforts to verify each of the materials they employ.

W. Powell 1998; Lori Rosenkopf and Michael L. Tushman 1998), and the role of science in technological search (Olav Sorenson and Lee Fleming 2004).

[4] See Stern (2004), Raymond Cypess (2003), and OECD (2001) for detailed discussion of the function, history, and policy analysis of BRCs.

[5] Even with recent advances in verification procedures, some researchers argue that a substantial fraction of currently circulated cell lines are still misidentified (Roderick A. F. MacLeod et al. 1999; John R. Masters 2002; Roland M. Nardone 2007; Rhitu Chatterjee 2007).

The problem of maintaining the integrity of research materials is not simply technological, but is driven by the economics of research incentives. Though a robust system for validating experimental research is collectively in the interest of all scientists, individual researchers have few incentives to invest in replication and validation. Indeed, researchers may find it worthwhile to limit scrutiny of their published results, at least in the short term. As the integrity of the scientific process is a public good, an institutional response is essential.

Several alternative institutional arrangements exist, including peer-to-peer networks, for-profit and proprietary culture collections, and biological resource centers. Peer-to-peer networks consist of informal exchanges among researchers and are dependent on research laboratories maintaining modest-sized culture collections and fulfilling requests for distribution on an ad hoc basis. In a pure peer-to-peer network, it may not be possible to require researchers to exchange materials, and initial discoverers may be reluctant to offer access to those whose experiments could undermine the value of the initial work (Eric G. Campbell et al. 2002; John P. Walsh, Charlene Cho, and Wesley M. Cohen 2005).[6] The potential for authentication problems is substantial in peer-to-peer networks, as labs rarely possess leading-edge verification tools and rely on (poorly paid) assistants to circulate research materials.[7] An alternative is proprietary collections, such as those maintained by major biopharmaceutical firms and for-profit biomaterials distribution firms. Not surprisingly, each type of collection "cherry picks" a narrow range of materials (the vast majority of which have already been accessioned at a major BRC) and focuses on materials with low storage costs and near-term commercial rewards.

BRCs, in contrast, pursue the objective of enhancing scientific research productivity by providing access to standardized biological materials. The World Federation of Culture Collections lists more than 550 of these "living libraries," whose members' collections exceed 1.4 million organisms (WFCC 2009).[8] Most countries have national collections that rely principally on government financing, helping to ensure that materials are accessioned based on their long-term scientific potential rather than near-term commercial concerns. Smaller collections then serve specialized research communities. The single largest BRC is the US-based American Type Culture Collection (ATCC, the "Library of Congress for biological materials"), which maintains a library of more than a million materials, distributes more than a quarter of a million materials annually, is supported by a mix of public and private funding, and is governed by a board that includes eminent life science researchers (ATCC 2009). Relative to alternative institutional arrangements, four distinctive

[6] Peer-to-peer transactions require researchers to contract with the developer of a particular biomaterial. In some cases, negotiations over access require the recipient to offer coauthorship, another incentive, or even a Materials Transfer Agreement or patent license. While such arrangements were rarely required for academic researchers during the bulk of our study period, the use of IP in academic science has become prevalent (and controversial) over the last decade (Michael A. Heller and Rebecca S. Eisenberg 1998; Mowery and Arvids A. Ziedonis 2007; Fiona Murray and Stern 2007; Cohen and Walsh 2008; Murray 2010; Murray et al. 2009; James Evans 2004).

[7] Informal brokers may emerge in peer-to-peer networks, facilitating transactions (Naomi R. Lamoreaux and Kenneth L Sokoloff 1999; Joshua S. Gans and Stern 2003). However, brokers are limited by the extent of their personal networks, and, since it is difficult to verify who is responsible when shared materials become contaminated, the potential for a purely reputation-based system may be limited.

[8] Life scientists and science policy analysts have emphasized the importance of BRCS in scientific progress and suggested that their importance has increased over the past 25 years (Hunter Cevera 1996; OECD 2001; David Smith 2003; Stern 2004).

attributes of BRCs may be associated with enhancing the accumulation of knowledge across research generations:

(i) *Certification:* Biomaterials authentication is one of the primary functions of BRCs. When accessioning materials into their collections (and periodically thereafter), BRCs subject materials to reviews and tests to verify their identity and biological viability. Relative to the peer-to-peer network, BRCs have mission-based incentives to establish a reputation for quality across a wide range of biological materials, are able to amortize the fixed costs of certification across multiple users of a given material, and invest in the specialized equipment and skills required for the certification of biomaterials. Of course, the returns to certification may vary: in particular, the value of certification may be particularly high for biomaterials initially disclosed in less prestigious journals, since the quality signal associated with those journals may be more variable.

(ii) *Independent and Open Access to Biological Materials:* BRCs ensure that their materials are equally accessible to all members of the scientific and technological community, thus encouraging *independent and open access* to the results of prior scientific research. While access through the peer-to-peer network is limited by the incentives of individual scientists to provide biomaterials access to potential scientific rivals, BRCs sever the direct tie between the researcher responsible for the initial discovery and those wanting to build upon the research. BRC collections reduce opportunities for hold-up through standardized Materials Transfer Agreements (MTAs). By facilitating the usage of materials by researchers in disparate scientific fields or at institutions that do not have access to a material through the peer-to-peer network, BRCs can expand the range of impact of a given scientific discovery.

(iii) *Preservation of Biological Materials:* Unlike the collections of individual researchers or for-profit organizations, BRCs are dedicated to the long-term maintenance of a broad range of materials whose value may not be initially apparent. BRCs have developed capabilities to enhance the value of materials over time and enable high-impact discoveries to be made many years after the initial discovery of a particular biomaterial. For example, *Thermus aquaticus* (Taq), a microorganism discovered in the hot springs of Yellowstone National Park in the late 1960s, is an *extremophile* that can sustain enzymatic reactions during rapid heating and cooling. While no practical benefit was seen at the time of its initial deposit at ATCC, its availability and preservation were fundamental in the development of biotechnology. More than 15 years after its discovery, a private sector researcher, Kary Mullis, was able to exploit the Taq extremophile in the development of polymerase chain reaction (PCR) to dramatically enhance the ability to replicate and sequence genetic material, earning Mullis the Nobel Prize in 1993 and Taq *Science*'s Molecule of the Year honor in 1989. Whereas individual researchers focus on maintaining only those materials required for their own research needs, and for-profit distributors focus on high-volume materials with low storage costs, BRCs' explicit objective of maintaining

an "option" on biomaterials leads to the active and careful archiving of a wide range of materials.

(iv) *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 that lower the costs of cumulative research. Relative to other organizational forms for preserving and circulating life science materials, BRCs maintain larger, more varied, and more balanced collections and reduce duplicative effort. As a result, BRCs are more likely to undertake the R&D and capital investments necessary to increase the quality and reduce the cost of accessing biological materials. For example, the size and breadth of their collections have enabled institutions such as the ATCC, DSMZ, the Coriell Institute, the Japan Collection of Microorganism, and the Jackson Laboratory to establish positions of global leadership in specific materials and collections, in authentication techniques, and in bioinformatics.

III. Identification Strategy

By ensuring the fidelity and lowering the costs of access to knowledge, research-enhancing institutions such as BRCs may influence the equilibrium rate and impact of a given discovery on subsequent research. Four central predictions stand out. First, a selection effect implies that, on average, knowledge associated with BRCs will be of higher intrinsic scientific value than knowledge that is available only through alternative institutions, such as the peer-to-peer network. Second, conditional on the intrinsic importance of a particular discovery, accession by a research-enhancing institution confers a positive marginal impact on subsequent knowledge diffusion. Furthermore, since research-enhancing institutions preserve access to knowledge for a longer time than alternative institutions, a preservation effect may arise in which the marginal impact persists (or grows) rather than erodes over time. Third, the marginal impact of a research-enhancing institution will be greater for knowledge associated with "poor" institutional environments. Finally, the extent of follow-on research induced by association with a research-enhancing institution will be greater among researchers and projects for which authentication and independent access are more valuable.

To evaluate these hypotheses, we must address a fundamental inference problem. For a given piece of knowledge within a given institutional environment, one cannot observe the counterfactual impact that such knowledge would have, had it been produced and diffused in an alternative institutional setting. From an experimental perspective, the econometrician would ideally assign discoveries randomly to distinct institutional environments and then compare the impact of different regimes on follow-on research use. While one cannot replicate this ideal experimental design, we develop an econometric strategy that takes advantage of exogenous institutional changes to isolate the marginal impact of an institution on knowledge accumulation from the effect of selection into that institution. Our approach exploits two key elements of our setting. 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.[9] Second, while initial publication often occurs within six months (or fewer) after initial journal submission, many BRC material deposits occur long after the publication date of the associated scientific research article. Moreover, in certain instances discussed in the next section, the act of deposit and its precise timing are arguably econometrically exogenous (and we can apply difference-in-differences techniques to test this assumption). Specifically, we observe several instances where principal investigators retire or change institutional affiliations, resulting in the transfer of special collections of materials from academic laboratories into a BRC.

Conditional on our assumption that the timing of the deposit is exogenous, the deposit lag allows us to estimate the impact of deposit on knowledge diffusion, measured as the change in the rate of citation (between the predeposit and post-deposit period) to the initial article by follow-on scientific research articles. We construct a dataset composed of scientific publications linked to (delayed) BRC deposits and control articles that are comparable to our treatment articles. Because we observe citations to a scientific publication both before and after BRC deposit (and because we are able to identify a counterfactual estimate of the citation rate that would have occurred if a BRC deposit had *not* occurred), we can identify the causal impact of BRC deposit on the pattern of citations to a scientific publication. Citations data take the form of count data that are skewed to the right and overdispersed relative to Poisson. Additionally, the rate of citation to a given piece of research will vary with the calendar year and the time elapsed since initial publication. In our regressions, we therefore employ a conditional negative binomial model with age and year fixed effects for citations produced per year for each scientific article in our dataset.[10,11]

To disentangle the treatment effect from the selection effect, we develop an initial estimator that identifies both the average differences between the treatment and control groups and the change in citations resulting from BRC deposit. This

[9] Most life sciences papers are short and focused, with few extraneous references beyond those directly affecting the described findings. Thus, the principal rationale for the inclusion of a citation for a BRC-linked paper is that the material is explicitly used in a follow-on experiment or the experiment is closely connected to the findings and knowledge linked to that specific material (i.e., life sciences citations are likely more informative than social science citations). More generally, the meaning and use of academic citations has become the subject of a large body of research, including the field of scientometrics (Eugene Garfield and Uriel H. Schoenbach 1956; Garfield 1979; Derek de Solla Price 1976; Loet Leydesdorff 2001). While recent papers suggest the potential for strategic and reputation-based citation (M. V. Simkin and V. P. Roychowdhury 2003), the focused nature of BRC-linked citations likely mitigates this concern.

[10] Panel data estimation of fixed-effects count data models must address several subtle issues, including the incidental parameters problem (Jerry Hausman, Bronwyn H. Hall, and Griliches 1984; Paul D. Allison and Richard P. Waterman 2002; William Greene 2004) and restrictions implied by distributional assumptions (Jeffrey M. Wooldridge 2002). We have experimented with both (i) conditional and dummy fixed effects estimators (trading off asymptotic consistency for small sample bias) and (ii) quasi-ML Poisson and negative binomial estimators (trading off robustness to specification error versus a more flexible distribution). Our results are based on the traditional conditional fixed-effects negative binomial estimator with bootstrapped standard errors; however, the key findings are consistent across these different procedures.

[11] When using a conditional fixed-effects estimator, one citation year and one age fixed effect are not separately identified (Hall, Jacques Mairesse, and Laure Turner 2007). Since the main effect that we are interested in is separable from these effects, the precise specification we employ to overcome this identification issue does not at all affect our estimate of the impact of BRC deposit on citations. In our estimation, we identify differences relative to age $= 0$, and relative to publication in years after 1975 (though, due to data limitations, we actually impose a single regressor on the years 1975–1979).

specification includes "article pair" effects that identify each matched treatment and control article, a dummy variable for all BRC-linked articles (identifying the selection effect), and a dummy variable for BRC-linked articles in the years after BRC accession (identifying the treatment effect). In subsequent regressions, we employ article fixed effects (conditional negative binomial fixed effects), thus identifying the treatment effect though not the selection effect. Building from these base specifications, we experiment with a range of related regressions that examine the robustness of the results to timing effects and heterogeneity in both the root and citing articles. We also take advantage of the structure of the data in order to identify the treatment effect using only the treated articles. We describe each of the specific estimating equations as we review specific findings in Section IV.

In addition to traditional concerns about interpreting citations (Griliches 1990; Keith Pavitt and Pari Patel 1988), we are careful to consider the possibility that substitution is biasing the results. For example, *citation substitution* may arise if materials deposits lead future researchers to cite BRC-linked articles rather than other articles that reflect the same knowledge, while *materials substitution* could arise if accession leads to an increase in citations to papers using the deposited material rather than to papers using substitute materials. Switching among close but imperfect substitutes (e.g., from a mutated version of a cell line that circulates within the scientific research network to the material included in a BRC deposit) might lead to a significant increase in citations without a significant increase in overall research productivity or quality.[12] For example, for very popular materials (such as HeLa), there may be several "independent" versions circulating within the scientific community. Our research design mitigates the possibility. By analyzing materials included in the "special collections," we focus on materials that are sufficiently specialized that there are few close substitutes (other than materials in the collection itself) and for which there was a low likelihood of a "secondary market." We nonetheless test for the possibility of substitution in our empirical analysis by examining whether exogenous deposits *negatively* affect citations to articles that are likely substitutes for BRC-deposited materials.

IV. Data

A. *Data Construction and Sources*

To conduct the empirical analysis, we focus on materials associated with a single institution, the American Type Culture Collection. Located in Manassas, Virginia, and founded in 1925, ATCC maintains the largest culture collection in the world (ATCC 2009). Although ATCC is unusually large, its preservation, certification, and distribution functions are similar to those of other large public culture collections. We take advantage of the characteristics of ATCC in order to address four key empirical challenges associated with implementing the difference-in-differences strategy we articulate above: (i) linking BRC deposits to research publications,

[12] It is also possible that there are multiple *identical* versions of a biological material maintained by different laboratories. Since these materials would be perfect substitutes from the perspective of cumulative knowledge production, strains that are identical to BRC deposits will be considered effectively part of the ATCC collection.

(ii) selecting a sample of publications that can be used to identify the marginal impact of BRCs, (iii) constructing a sample of control articles, and (iv) accounting for ambiguity in the date on which BRC deposits are available for follow-on research.

We address the first challenge by taking advantage of the reference information maintained by ATCC on all materials deposited in its collections. For each material, ATCC documents the name of the original depositor, date of deposit, and key scientific information associated with the deposit, including the key research article that employs or characterizes the material.[13]

To overcome the second challenge, we take advantage of shocks that led to the mass transfer of three *special collections* into ATCC from collections previously circulated via the peer-to-peer network. These transfers occurred when scientists who maintained collections within the peer-to-peer network moved or faced an institutional funding limitation unrelated to that specific collection. The first set of materials is drawn from the Tumor Immunology Bank (TIB), which was accessioned into ATCC beginning in 1982 due to funding pressures at the Salk Institute, where it had been previously maintained. The second special collection is the Human Tumor Bank (HTB), which had been operated by researchers at Sloan-Kettering until institutionwide funding considerations led to its wholesale transfer beginning in 1981. The third special collection, the Gazdar Collection, was transferred into ATCC beginning in 1994 when Dr. Adi Gazdar left his position as Head of Tumor Cell Biology at the National Cancer Institute, and, along with his collaborator Dr. John Minna, moved to UT-Southwestern. It is important to note that the materials in each collection were (i) publicly available as part of the special collection prior to the transfer to ATCC, (ii) unavailable from proprietary vendors during the sample period, and (iii) unencumbered by formal intellectual property claims such as patents or MTAs. Together, there are 72 articles matched to materials in the TIB collection, 30 from the HTV collection, and 6 from the Gazdar collection.

We additionally identify a set of control articles for each BRC-affiliated article, using the "most-related" article in the same volume of the journal in which the BRC-linked article was published. We identify most-related articles based on a search algorithm developed by the National Library of Medicine (NLM). The NLM algorithm generates similarity rankings based on the extent to which articles in the PUBMED database share terms in their title, abstract, and Medical Subject Headings (MeSH). From the set of articles identified by the NLM algorithm as related to the focal article, we select the most-related article published in the same journal and publication year.[14]

The fourth challenge is to account for ambiguity in the date on which BRC deposits are available for access by other researchers. Some members of the research community become informed about collections transfer through informal communications and formal announcements prior to the official accession date. At the same

[13] Historically, ATCC published its catalogs in print form. Currently, ATCC maintains its catalog online at www.ATCC.org. In cases in which multiple publications are relevant for a particular material, we use the first article listed, as ATCC scientific and information technology staff report that this is the article most closely associated with the initial use of the biological material.

[14] In cases in which no article in the same volume of the journal qualifies as sufficiently related according to the NLM algorithm, we use the article that immediately precedes the BRC-linked article in the specific year and issue in which the BRC-linked article was published as the control. For example, if a BRC-linked publication were the third article in the June 14, 1986, issue of *Cell*, the control article would be the second article in that same issue.

time, because of the rigorous procedures used to accession materials, some materials in the HTB and TIB collections took 24 months to be officially declared available from ATCC. We explicitly account for this transition period by incorporating a "transfer window," including the year before, the year of, and the year following the official accession date. By including this window, our analysis focuses on how the pattern of citation changes from a period prior to the deposit announcement to the period subsequent to its availability through a BRC. We also compile detailed bibliometric information, including annual citation counts and bibliometric details of cited and citing articles from the Institute for Scientific Information (ISI) Science Citation Index Expanded (SCI) database.[15]

B. *Summary Statistics*

Our core dataset consists of 108 BRC-linked articles and 108 associated control articles. We refer to these articles as "root articles" to distinguish them from the "citing articles" that reference them. Table 1 provides variable names and definitions, and Table 2A reports summary statistics. We track citations to each root article from the year of its publication (mean *PUBLICATION YEAR* = 1979.4), yielding 4,857 article-year observations. The majority of BRC-linked articles were deposited in the early 1980s, although the articles associated with the Gazdar collection were published in the early 1990s. Root articles in the sample are predominantly associated with US-based authors (76 percent); 15 percent are associated with the top 50 most research-intensive US universities; and slightly more than half (56 percent) appear in journals with an ISI impact factor greater than 25.[16] Our sample includes citations received by root articles between 1970 (the earliest publication year) and 2001, and the citation-years have an average *AGE* of 11.3 years. The key dependent variable in our analysis is *FORWARD CITATIONS*, which measures the number of citations received by a root article in a given year. Because publications associated with BRC deposits (and their associated control articles) tend to appear in top-tier journals, such as *Science*, *Nature*, and *Cell*, the average number of forward citations is higher than would be expected for a randomly chosen life sciences article. The average number of annual *FORWARD CITATIONS* in our sample is 7.28, the cumulative number of citations by 2001 is 91.7, and the distribution is, not surprisingly, skewed.

To examine heterogeneity in the treatment effect, we have also gathered detailed bibliometric information from the set of citing articles. We construct several measures of the number of citations that a root article receives from specific types of articles, including annual citations from papers with US-based authors (mean = 2.6), annual citations from articles associated with a top 50 US university (mean = 1.0), annual citations from articles appearing in top journals (mean = 3.4), and annual citations from articles with a single ISI subject category (mean = 4.1) as opposed

[15] The SCI has been widely used in economics, sociology, and management research, as well as in bibliometric studies, to quantify scientists' research output, measure research collaboration, and track the diffusion of science—prominent examples include Sharon G. Levin and Paula E. Stephan (1991), James D. Adams and Griliches (1998), Iain M. Cockburn and Henderson (1998), and Lynne G. Zucker, Michael R. Darby, and Marilynn B. Brewer (1998).

[16] The average numbers of authors per article is 5.0, pages is 6.6, backward citations is 31.9, and BRC material price is $223.

Table 1—Variables and Definitions

Variable	Definition	Source
Citation characteristics		
FORWARD CITATIONS$_{jt}$	Number of *FORWARD CITATIONS* to article *j* in year *t*	Science citation index (SCI)
CUMULATIVE CITATIONS$_{jt}$	Number 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))	ATCC
	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 it 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.	
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
US AUTHOR	Dummy variable equal to 1 if reprint author (corresponding author) associated with an institution located in the United States; 0 otherwise	SCI; author verification
TOP 50 UNIVERSITY	Dummy variable equal to 1 if reprint author (corresponding author) is associated with an institution that appears in the US top 50 according to the Center for Measuring University Performance (Arizona State University) 2006 Annual Report of university research rankings	CMUP (ASU)
TOP JOURNAL	Dummy variable equal to 1 if article appears in a journal with ISI Journal Impact Factor greater than 25.	SCI; author verification
Citing article characteristics		
CITES FROM US RP AUTHOR ARTICLE	Count of citations from reprint author (corresponding author) associated with an institution located in the United States	SCI; author verification
CITES FROM TOP JOURNAL	Dummy variable equal to 1 if article appears in a journal with ISI Journal Impact Factor greater than 25.	SCI; author verification
CITES FROM ARTICLE WITH SINGLE SUBJECT CATEGORY	Count of citing articles associated with only a single ISI scientific subject category, based on the ISI broad subject category classification scheme.	SCI
CITES FROM ARTICLE WITH MULTIPLE SUBJECT CATEGORIES	Count of citing articles associated with more than one ISI scientific subject category, based on the ISI broad subject category classification scheme.	SCI

TABLE 2A—MEANS AND STANDARD DEVIATIONS

Variable	Mean	SD	Min	Max
Article characteristics ($n = 216$ articles)				
BRC ARTICLE	0.50	0.50	0	1
PUBLICATION YEAR	1979.40	4.54	1970	1992
*DEPOSIT YEAR**	1983. 63	3.47	1981	1994
US AUTHOR	0.76	0.43	0	1
TOP 50 UNIVERSITY AUTHOR	0.15	0.36	0	1
TOP JOURNAL	0.56	0.50	0	1
Article-year characteristics ($n = 4{,}857$ article*year observations)				
YEAR	1989.79	7.23	1970	2001
AGE	11.27	7.23	0	31
FORWARD CITATIONS	7.28	15.73	0	186
CUMULATIVE CITATIONS	91.67	178.86	0	2333
Forward citations received from papers with				
US AUTHOR	2.60	5.87	0	59
TOP 50 UNIVERSITY AUTHOR	0.99	2.50	0	33
TOP JOURNAL	3.37	8.02	0	99
SINGLE SUBJECT CATEGORY	4.10	10.78	0	138
UNIQUE NEW JOURNALS[a]	2.88	4.60	0	65
UNIQUE NEW INSTITUTIONS[a]	5.56	10.23	0	143
UNIQUE NEW COUNTRIES[a]	0.79	1.41	0	16

Notes: * *DEPOSIT YEAR* data only for BRC-linked articles (108 articles; 2,441 article-years)

[a] *CITATIONS BY UNIQUE NEW JOURNALS*, *INSTITUTIONS*, and *COUNTRIES* refer to citations in a particular year from journals, institutions, and countries that had not cited the root article in previous years. For example, if an article were to receive ten citations in its first year after publication, all of which appeared in *Science*, and two citations in its second year after publication, one that appeared in *Science* and the other that appeared in *Nature*, then *CITATIONS FROM UNIQUE NEW JOURNALS* would equal one in the first year (since all publications appeared in the same journal, *Science*) and one in the second year (although two separate journals cited the root article in that year, only the citation in *Nature* is unique, as a citing article had appeared in *Science* in the previous year).

to multiple subject categories.[17] We also construct measures capturing the number of citations received from articles with identifiers that are new to the set of citations associated with a given root article. These measures are intended to reflect increases in the "breadth" of the research community drawing on the knowledge in a particular root article. Specifically, we construct three variables: *CITATIONS BY UNIQUE NEW JOURNALS* (mean = 2.9), *CITATIONS BY UNIQUE NEW INSTITUTIONS* (mean = 5.6), and *CITATIONS BY UNIQUE NEW COUNTRIES* (mean = 0.8). Each of these measures refers to citations in a particular year from journals, institutions, and countries, respectively, that had not yet cited the root article in previous years.[18]

Table 2B compares key characteristics of the BRC-linked articles to those of the control sample. Articles associated with BRC deposits receive greater than 220 percent more citations than *Most-Related Article* controls, even though both control groups appear in the same journal, went through the same review process, and are

[17] The ISI has developed a scheme for classifying academic research into detailed scientific subject categories, including "Biochemical Research Methods," "Cell Biology," and "Oncology." The SCI includes a field identifying the subject category or categories into which journals and papers have been classified. Journals and papers that cross scientific areas may be assigned multiple subject categories. Papers in our sample receive a minimum of one and a maximum of five subject categories.

[18] For example, if an article were to receive ten citations in its first year after publication, all of which appeared in *Science*, and two citations in its second year after publication, one that appeared in *Science* and the other that appeared in *Nature*, then *CITATIONS FROM UNIQUE NEW JOURNALS* would equal one in the first year (since all publications appeared in the same journal, *Science*) and one in the second year (although two separate journals cited the root article in that year, only the citation in *Nature* is novel, as a citing article had appeared in *Science* in the previous year).

TABLE 2B—MEANS AND STANDARD DEVIATIONS, BY CONTROL GROUP

	Treatment articles: Articles associated with ATCC deposits	Control articles: Most-related article control
Number of papers	108	108
Paper-years (max)	2,441	2,439
FORWARD CITATIONS	11.13	3.47
	(19.64)	(9.01)
CUMULATIVE CITATIONS	137.57	45.86
	(230.22)	(82.46)
PUBLICATION YEAR	1979.40	1979.40
	(4.55)	(4.55)

matched closely by subject area. Figure 1 portrays the disparity between these groups over time, comparing average citations by article age. For each sample, the average number of citations increases over the first few years, peaking around the third or fourth year after publication.

V. Empirical Results

A. *Baseline Analyses*

Our baseline analysis begins with estimations that identify the effect of selection into the ATCC collection separately from the marginal impact of ATCC deposit on subsequent citation. Thereafter, we focus on identifying the magnitude and nature of the marginal effect. To disentangle the marginal impact of ATCC deposit from the selection effect, we develop a difference-in-differences estimator that identifies the average differences in citations received between treatment and control articles (pairing each article in the treatment group with a "similar" article in a control group) and the change in citations that results from BRC deposit for those articles ultimately accessioned into a BRC collection. Specifically, we estimate variations of

$$
\begin{aligned}
(1)\qquad & \textit{FORWARD CITATIONS}_{i,j,t} \\
& = f(\varepsilon_{i,j,t}; \alpha_j + \beta_t + \delta_{t-\textit{pubyear}} + \phi \textit{BRC-ARTICLE}_i \\
& \quad + \psi_{\textit{WINDOW}} \textit{BRC-ARTICLE} \times \textit{WINDOW PERIOD}_{i,t} \\
& \quad + \psi \textit{BRC-ARTICLE} \times \textit{POST-DEPOSIT}_{i,t}),
\end{aligned}
$$

where α_j is a fixed effect for each pair of a treatment article and control article, β_t is a year effect, $\delta_{t-\textit{pubyear}}$ captures the age of the article, and *BRC-ARTICLE* is a dummy variable equal to one for those articles linked at some point to a BRC. BRC-*ARTICLE* × *WINDOW PERIOD* is a dummy variable equal to one during the

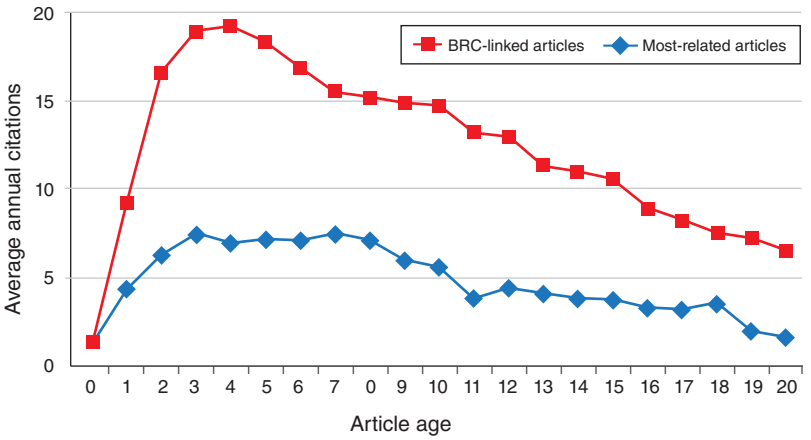

FIGURE 1. AVERAGE ANNUAL CITATIONS BY AGE, BRC VERSUS CONTROL ARTICLES

year immediately prior to, the year of, and the year immediately after the accession of a material into the BRC; this accounts for "announcement effects" and for potential lags in availability of materials. *BRC-ARTICLE* × *POST-DEPOSIT* is a dummy variable equal to one only in those years after the material linked to the article has been accessioned into and is available from a BRC.

The first three columns of Table 3 report results based on equation (1). Columns 3–1 and 3–2 begin with OLS specifications using ln(*FORWARD CITATIONS*) as the dependent variable. Column 3–1 omits *YEAR* fixed effects and fixed effects for each treatment-control pair. The results are similar: the marginal impact of BRC deposit (controlling for the selection effect) is estimated to be in excess of a 50 percent boost to the citation rate. Moreover, ϕ suggests that articles that are ultimately linked to BRC deposits are associated with a 50 percent higher citation rate relative to the controls (i.e., the selection effect in this sample is large and positive). Also, the marginal impact of BRC deposit begins to manifest itself during the window period (with an estimated 33 percent boost), and both the year and article age fixed effects are jointly significant (although the interpretation of such a test is subtle (Hall, Mairesse, and Turner 2007; Aditi Mehta, Marc Rysman, and Tim Simcoe 2010)).

OLS is useful as a preliminary exercise, but inappropriate for inference, as citation data are composed of highly skewed count data. We therefore employ a conditional fixed effects negative binomial estimator for the remaining specifications. We report in brackets the coefficients for these models as incidence-rate ratios (a coefficient equal to one implies no effect on *FORWARD CITATIONS*, whereas a coefficient equal to 1.50 implies a 50 percent boost to *FORWARD CITATIONS*).[19] The first of these specifications, 3–3, presents a useful comparison to 3–2. We can easily reject

[19] All models include block bootstrapped standard errors, clustered either by article pairs or article dummies, depending on the set of fixed effects included in the specification (Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan 2004; James G. MacKinnon 2002). We do not report the significance of tests of joint restrictions on the article family or article fixed effects, as these are not computed in conditional fixed-effects models.

Table 3—Baseline Specifications

	OLS (Robust SEs, adjusted for clustering by article group, are reported in parentheses) Dep Var = ln(*FORWARD CITATIONS*)		*CONDITIONAL FIXED EFFECTS NEGATIVE BINOMIAL** [Incidence-rate ratios in brackets in top line] Estimated coefficients in 2nd line. (Block bootstrapped SEs reported in parentheses) Dep Var = *FORWARD CITATIONS*	
	Base model: BRC effect with age FEs only (3-1)	Base model, with article family and year FEs (3-2)	Baseline count model (3-3)	Baseline diff-in-diffs specification (3-4)
Article characteristics				
BRC-ARTICLE			[2.121]	
	0.497	0.501	0.752	
	(0.156)***	(0.132)***	(0.397)***	
BRC-ARTICLE, WINDOW PERIOD			[1.422]	[1.759]
	0.332	0.385	0.352	0.565
	(0.125)***	(0.106)***	(0.234)**	(0.247)***
BRC-ARTICLE, POST-DEPOSIT			[1.713]	[2.248]
	0.536	0.535	0.538	0.810
	(0.177)***	(0.142)***	(0.348)***	(0.360)***
Control variables				
Parametric Restrictions				
Age FEs = 0	Sig.	Sig.	Sig.	Sig.
Article pair FEs = 0		Sig.		
Year FEs = 0[a]		Sig.	Sig.	Sig.
Constant	0.138	2.213		
	(0.087)	(0.111)***		
Observations	4,857	4,857	4,753	4,729
R^2	0.24	0.54		
Log likelihood			−10,759.18	−9,632.40
Number of article pairs			106	
Number of articles				211

[a] Year FEs included for 1980–2001; 1970–1974 and 1975–1979 grouped.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

the null of no selection and no marginal effect. Indeed, the estimated coefficients are larger than those associated with the OLS specifications and suggest the practical significance of the treatment effect: forward citation rates are estimated to increase more than 70 percent *after* BRC deposit. Moreover, BRC-linked articles receive 112 percent more citations annually than matched control articles, implying that articles associated with the special collections were of greater intrinsic scientific importance than those in the control sample.[20]

[20] It is useful to note that this estimate of the selection effect is specific to this sample and empirical design and does *not* serve as an estimate of the average selectivity of BRC-linked articles: our sample of treatment articles is not a random sample of BRC-linked articles (we chose those articles that were subject to an exogenous deposit), and the control articles are not a random sample of life sciences articles (we chose those articles to be close matches to the treatment articles).

To identify the marginal effect of deposit on subsequent citations more precisely, we modify our initial specification to account for heterogeneity across matched article pairs. In doing so, we incorporate article-specific fixed effects (γ_i) into equation (1), resulting in

$$(2)\quad FORWARD\,CITATIONS_{i,t} = f(\varepsilon_{i,t};\gamma_i + \beta_t + \delta_{t-pubyear} + \psi_{WINDOW} BRC\text{-}ARTICLE \times WINDOW\,PERIOD_{i,t} + \psi BRC\text{-}ARTICLE \times POST\text{-}DEPOSIT_{i,t}).$$

This specification tests for the impact of research-enhancing institutions by calculating how the citation rate for a publication *changes* after BRC deposit, accounting for fixed differences in the citation rate across articles and relative to the nonparametric trend in citation rates for articles with similar characteristics. Except where noted, all remaining specifications include (conditional) article fixed effects to account fully for heterogeneity in the underlying quality of individual articles. With the control articles helping to identify the citation year and article age effects, ψ is identified from the change in citations (relative to expectations) after the associated biomaterial is accessioned (and after the deposit window has elapsed).

We implement equation (2) in column 3–4. The results suggest that BRC-linked articles receive a 125 percent citation boost after BRC accession, controlling for article, age, and year-specific effects.[21] Of course, the interpretation of this estimate depends on the extent to which the coefficient reflects the marginal treatment impact of BRC deposit, as opposed to spurious correlation. We therefore test our key identification assumptions in Table 4. We first examine whether the results in Table 3 are simply the result of a different citation age profile for BRC-linked articles compared to the controls.[22] For example, BRC-linked articles may have inherently longer-lived citation profiles, which would result in an upward bias on the estimate of *BRC-ARTICLE* × *POST-DEPOSIT*. We address this possibility in two distinct ways. First, in column 4–1, we include a separate linear time trend for BRC-linked articles; the coefficient on *BRC-ARTICLE* × *AGE* is positive but insignificant. While ψ declines relative to 3–4, the effect of BRC deposit remains statistically and quantitatively significant and similar in magnitude to 3–3. In 4–2, we consider the citation age profile more precisely by also accounting for the preservation hypothesis—the idea that the impact of BRC deposit may grow with the time elapsed from the deposit date. When we also include the regressor, *YEARS SINCE BRC DEPOSIT*, we cannot empirically disentangle the BRC-specific age trend from the trend that may arise after BRC deposit. This is not surprising,

[21] These overall findings are robust across a wide range of alternative subsamples and control groups, including the exclusion or inclusion of any special collection (TIB, HTB, and Gazdar), a control sample composed exclusively of nearest neighbor controls (the articles immediately preceding treatment articles in the journal and volume in which they appear), or a sample that includes only a treatment and control article when a most related article control is available (Furman and Stern 2006).

[22] In this analysis, we continue to include the control articles to identify the impact of year effects and nonparametrically estimate the shape of the age profile, while also including a separate BRC-linked age trend:

$$FORWARD\,CITATIONS_{i,t} = f(\varepsilon_{i,t};\gamma_i + \beta_t + \delta^0_{t-pubyear(i)} + \delta^1 BRC\text{-}ARTICLE \times AGE_i + \psi BRC\text{-}ARTICLE \times POST\text{-}DEPOSIT_{i,t}).$$

TABLE 4—ACCOUNTING FOR THE AGE PROFILE OF BRC-LINKED ARTICLES

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL [Incidence-rate ratios in brackets in top line] Estimated coefficients in second line (Block bootstrapped SEs reported in parentheses) Dep Var = *FORWARD CITATIONS*			
	Interacting BRC-article × age (4-1)	Accounting for BRC-article age and time since deposit (4-2)	Identification based only on variation within BRC-linked sample (with grouped year FEs) (4-3)	Identification based only on variation within BRC-linked sample (with polynomial expansions for year and BRC-age) (4-4)
Article characteristics				
BRC-ARTICLE, WINDOW PERIOD	[1.515] 0.415 (0.302)**	[1.514] 0.415 (0.361)*		
BRC-ARTICLE, POST DEPOSIT	[1.677] 0.517 (0.438)**	[1.676 0.516 (0.474)*	[1.633] 0.490 (0.351)**	[1.576] 0.455 (0.312)**
BRC-ARTICLE × AGE	[1.028] 0.027 (0.018)	[1.028] 0.028 (0.038)		
YEARS SINCE BRC-DEPOSIT		[1.000] 0.000 (0.040)		
Control variables				
Parametric restrictions				
Age FEs	Sig.	Sig.		
Calendar year effects via single-year dummies[a]	Sig.	Sig.		
Calendar year effects via five-year dummies			Sig.	
Year				[1.130] 0.122 (0.053)***
Year-squared				[0.997] −0.003 (0.001)***
Age			[0.991] −0.009 (0.027)	[0.995] −0.005 (0.036)
Age-squared			[0.998] −0.002 (0.001)**	0.997 −0.002 (0.001)
Observations	4,729	4,729	2,041	2,041
Log likelihood	−9,620.40	−9,620.40	−5,124.31	−5,118.40
Number of groups	211	211	105	105

[a] Year FEs included for 1980–2001; 1970–1974 and 1975–1979 grouped.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

as it is difficult to infer evidence for the preservation hypothesis in the context of a difference-in-differences estimator in which there is a limited number of treatment effects and we are separately allowing for a separate treatment-group time trend.

Thus far, our analyses have incorporated both treatment and control articles, even though our identification approach relies principally on differences in

citations received before and after BRC accession. Because article age at time of deposit and year of deposit vary (at least to a certain extent), it is possible to identify the impact of BRC deposit solely from the set of BRC-linked articles (i.e., excluding a control group). This constitutes a powerful alternative to a traditional difference-in-differences approach, as it directly addresses the problem of constructing a synthetic control group for heterogeneous knowledge outputs, such as scientific publications (or patents, in other applications). If there were sufficient variation in the time from publication to deposit and in the calendar year of deposit, we could, in principle, incorporate a complete set of publication age and calendar year fixed effects in our analysis. Unfortunately, the HTB and TIB special collections were accessioned at relatively similar times (1980–1983); thus, the structure of the data in this paper is not rich enough to identify *BRC-ARTICLE* × *POST-DEPOSIT* under the most flexible specification. As a consequence, our estimates based only on the treated sample condition on article-specific fixed effects, article age effects, and citation year effects (through polynomials in article age and citation year):

$$
\begin{aligned}
(3)\quad & \textit{FORWARD CITATIONS}_{i,t} \\
& = f(\varepsilon_{i,j,t};\gamma_i + g^0(t;\beta) + g^1(t - \textit{pubyear}(i);\delta) \\
& \quad + \psi_{\textit{WINDOW}}\textit{BRC-ARTICLE} \times \textit{WINDOW PERIOD}_{i,t} \\
& \quad + \psi\textit{BRC-ARTICLE} \times \textit{POST-DEPOSIT}_{i,t}).
\end{aligned}
$$

We implement this specification in the final two columns of Table 4, including five-year grouped calendar year effects and linear and quadratic article age effects in 4–3 and a linear and quadratic term for both calendar year and article age in 4–4. In each of these specifications, the postdeposit treatment effect remains statistically significant and of a magnitude similar to 4–1. By excluding the control sample and identifying the marginal impact of BRCs from variation in the treatment sample, these results reinforce our earlier findings without making assumptions about the quality of the match between the treatment and control samples.

So far, our analysis has assumed that the timing of BRC deposit is exogenous. If BRC-linked articles experience a significant increase in forward citations in the years prior to accession, this would imply that the measured postdeposit effect is confounded with a predeposit trend, undermining our interpretation of ψ as a treatment effect. To examine this, we implement a specification similar to 3–4 but include dummy variables for each year preceding and following BRC deposit (along with complete article, age, and calendar year fixed effects). Figure 2 plots each of these estimates (in terms of the incidence-rate ratio minus one, where all effects are computed relative to the window period), along with upper and lower bounds for 95 percent confidence intervals. Two findings stand out. First, the predeposit citation pattern does not suggest a clear upward trend in the nine years prior to accession. While there is a slight uptick in forward citations in years two and three before the window period, this effect is noisy and sensitive to the

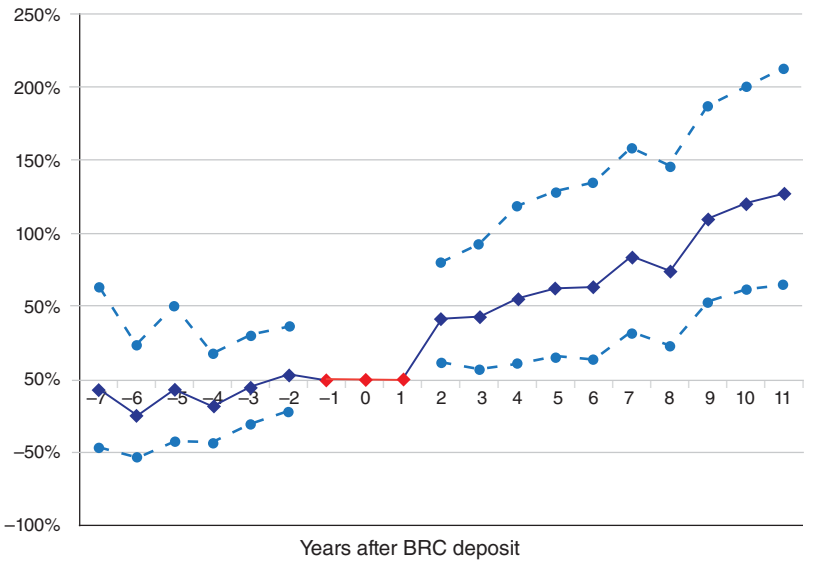

FIGURE 2. PRE- AND POSTDEPOSIT EFFECTS ON *FORWARD CITATIONS*

Notes: Figure plots year-by-year pre-deposit and post-deposit boosts (or decrements) to citations, computed from negative binomial regressions with dummy variables for each year preceding and following BRC-deposit (along with article, age, and calendar year fixed effects). The data represent each of the estimated pre- and post-deposit year effects, equal to the estimated incidence-rate ratio minus one, where all effects are computed (relative to the window period), along with upper and lower bounds for 95 percent confidence intervals.

estimation technique.[23] We cannot reject the hypothesis that all predeposit coefficients are equal. Furthermore, the sizable and near continuous increase in the citation boost in the years following deposit is consistent with BRC accession's having a significant marginal impact on *FORWARD CITATIONS*. While BRC-affiliated articles experience only a 40 percent citation boost in the years immediately after accession, this effect increases to over 125 percent by ten years after deposit. While this effect captures both the preservation effect and the potential for a positive trend in citations for BRC-affiliated articles (as captured in our discussion of 4–2), the evidence does suggest that the influence of BRC deposit does not decline over time. Whereas most research is used as an input in follow-on research for only a few years following publication, BRC-linked knowledge is "forgotten" at a much lower rate.

Our citation-based analytic approach also assumes that ψ reflects real changes in follow-on research behavior, rather than *citation substitution* or *materials substitution*. Although it is difficult to test for these practices directly, we investigate a straightforward implication of this possibility by examining whether exogenous deposits *negatively* affect citations to articles that are likely substitutes for BRC-deposited materials. Our approach is based on the idea that related articles by the authors of BRC-linked articles constitute a key set of potential substitute articles and identify a key set of potential substitute materials. If citation substitution were

[23] The predeposit trend has no upward trend in a specification with dummy fixed effects as opposed to conditional effects.

Table 5—Exploring Substitution between Articles

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL [Incidence-rate ratios in brackets in top line] Estimated coefficients in second line (Block bootstrapped SEs reported in parentheses) Dep Var = *FORWARD CITATIONS TO MOST-RELATED-OWN ARTICLE*
Article characteristics	
BRC-ARTICLE, WINDOW PERIOD	[1.656] 0.504 (0.261)***
BRC-ARTICLE, POST DEPOSIT	[1.748] 0.558 (0.318)***
Control variables	
Age FEs	Sig.
Year FEs	Sig.
Observations	4197
Log likelihood	−7,550.19

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

to occur, we would expect that listing an article in the ATCC catalog would lead to a shift in citation away from other articles associated with that material and to the article listed by ATCC. In particular, citation substitution would yield a relative decline in citations to other articles by the same authors, as citers standardize on the paper listed in the ATCC catalog. Similarly, materials substitution would lead follow-on researchers to choose ATCC-affiliated materials at the expense of other, similar materials. Since researchers often develop multiple versions of the materials they work with, one set of materials likely to be the close substitutes for a deposited material are those characterized in related articles by the same researcher.[24]

To test for these forms of substitution, we develop a sample of *Most Related Own Articles*. To assemble these data, we use the NLM algorithm to identify the top 40 most-related articles for each root article and then define the most related own article as the highest-ranked article in this set that includes the root article's last author, first author, or, if neither of these, multiple middle authors. We investigate the prospect of citation substitution and materials substitution in Table 5 by examining whether citations to *Most Related Own Articles* decline following BRC deposit, as would be expected if these forms of substitution were to occur. The results suggest the opposite: *Most Related Own Articles* experience a statistically and economically

[24] To evaluate the relative salience of citation substitution and materials substitution, we evaluated 30 most related articles by hand. We found that the vast majority (more than two-thirds) pertained to related research by the same authors, suggesting that BRC deposit may amplify the researcher's broader research agenda. There were a small number of articles that employed the same material as the one deposited into the BRC enhances; in each of these cases, the most related article used the BRC material in combination with other biological materials or discussed a particular experimental finding rather than the simple characterization of the material. In other words, the most related own articles sample is, indeed, a reasonable sample in which to test for some form of materials substitution, and our finding that the citations to these articles increases is evidence that substitutions are not a first-order driver of our main findings.

Table 6—Heterogeneity in Treatment Effects Across Root Articles

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL [Incidence-rate ratios in brackets in top line] Estimated coefficients in second line (Block bootstrapped SEs reported in parentheses) Dep Var = *FORWARD CITATIONS*		
	Post-deposit effects for papers outside and inside top journal set (6–1)	Post-deposit effects for papers generated by authors outside and inside top 50 universities (6–2)	Post-deposit effects for papers classified according to predeposit levels of citation (using quartiles) (6–3)
Article characteristics			
BRC-ARTICLE, WINDOW PERIOD	[1.209] 0.190 (0.107)***	[1.211] 0.191 0.124***	[1.169] 0.156 0.132***
BRC-Article, Post-Deposit type			
BRC-ARTICLE IN TOP JOURNAL	[1.708] 0.535 (0.238)**		
BRC-ARTICLE NOT IN TOP JOURNAL	[2.155] 0.768 (0.341)***		
BRC-ARTICLE FROM TOP 50 UNIVERSITY		[1.793] 0.584 0.349***	
BRC-ARTICLE NOT FROM TOP 50 UNIVERSITY		[1.870] 0.626 0.208***	
BRC-ARTICLE IN LOWEST CITATION QUARTILE AT TIME OF DEPOSIT (Q1)			[1.812] 0.594 0.365***
BRC-ARTICLE IN SECOND LOWEST CITATION QUARTILE AT TIME OF DEPOSIT (Q2)			[2.431] 0.888 0.553***
BRC-ARTICLE IN SECOND HIGHEST CITATION QUARTILE AT TIME OF DEPOSIT (Q3)			[2.006] 0.696 0.296***
BRC-ARTICLE IN HIGHEST CITATION QUARTILE AT TIME OF DEPOSIT (Q4)			[1.489] 0.398 0.250***
CONTROL VARIABLES			
Age FEs	Sig.	Sig.	Sig.
Year FEs	Sig.	Sig.	Sig.
Observations	4,860	4,860	4,860
Number of groups	215	215	215
Log likelihood	−9,911.76	−9,919.14	−9,905.89

Notes: Tests of joint restrictions:

(6-1): β(Post-Deposit BRC-Article in Top Journal) = β(Post-Deposit BRC-Article in Not Top Journal):
$\chi^2(1) = 2.64$; $\Pr > \chi^2 = 0.10$

(6-2): β(Post-Deposit BRC-Article from Top50 University) = β(Post-Deposit BRC-Article not from Top50 University):
$\chi^2(1) = 0.06$; $\Pr > \chi^2 = 0.81$

(6-3): β(Post-Deposit BRC-Article in Lowest Citation Quartile) = β(Post-Deposit BRC-Article in Highest Citation Quartile)
$\chi^2(1) = 0.67$; $\Pr > \chi^2 = 0.4145$

β(Post-Deposit BRC-Article in Lowest Citation Quartile) = β(Post-Deposit BRC-Article in second Highest Citation Quartile)
$\chi^2(1) = 0.25$; $\Pr > \chi^2 = 0.6183$

β(Post-Deposit BRC-Article in Lowest Citation Quartile) = β(Post-Deposit BRC-Article in second Lowest Citation Quartile)
$\chi^2(1) = 1.23$; $\Pr > \chi^2 = 0.2678$

β(Post-Deposit BRC-Article in second Lowest Citation Quartile) = β(Post-Deposit BRC-Article in Highest Citation Quartile)
$\chi^2(1) = 4.16$; $\Pr > \chi^2 = 0.0414$

β(Post-Deposit BRC-Article in second Lowest Citation Quartile) = β(Post-Deposit BRC-Article in second Highest Citation Quartile)
$\chi^2(1) = 0.75$; $\Pr > \chi^2 = 0.3875$

β(Post-Deposit BRC-Article in second Highest Citation Quartile) = β(Post-Deposit BRC-Article in Highest Citation Quartile)
$\chi^2(1) = 3.33$; $\Pr > \chi^2 = 0.0682$

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

significant increase in citations after the underlying biological material is accessioned into a BRC. Overall, the result is smaller in magnitude than the baseline specification using the root articles (as expected), and allows us to reject the hypothesis that these particular forms of substitution are driving the baseline results.

B. *Drivers of the Marginal Impact of BRC Deposit*

We now turn to a more detailed investigation of the sources of the marginal citations arising from BRC deposit. We begin in Table 6 by evaluating heterogeneity in the BRC treatment effect across different types of root articles. Consistent with the theoretical model of Mukherjee and Stern (2009), we anticipate that BRC deposit will have a higher impact for articles published in journals where the "quality" signal is more ambiguous. The key coefficients in our analyses are those estimated as interaction effects between (*BRC-ARTICLE*, *POST-DEPOSIT*) and various types of BRC-linked root article types. We continue to conduct conditional fixed-effects negative binomial regressions, thus suppressing the direct calculation of selection effects. In 6–1, we estimate whether the marginal impact of BRC deposit differs for root articles associated with top-tier journals. Following BRC deposit, we find a 70 percent boost in citations to BRC-linked articles published in top journals; in comparison, BRC-linked articles published in non-elite journals experienced a 115.5 percent increase in citations. The difference in these coefficients is significant at the 10 percent level, implying that the marginal impact of BRC deposit is higher for articles published *outside* of top-tier journals relative to those published in top-tier journals. This result is consistent with an interpretation in which more selective journals have a higher "bar" for the underlying quality and reproducibility of the experiment than less selective journals. As an additional check, we also test whether the quality of the university affiliation has a significant impact on the citation boost by comparing the impact of BRC deposit on publications from reprint authors in top 50 universities versus others.[25] In contrast to 6–1, we find no significant difference in the impact of university affiliation. This is, perhaps, not surprising, as the mechanism by which university quality may have a differential effect on citations is not as clear as the case for journal quality.

In a further examination of potential certification benefits, we investigate whether the impact of BRC deposit depends on an article's level of predeposit citation. To do so, we run a first-stage OLS regression of the number of cumulative citations *at the time of deposit* as a function of calendar year fixed effects and fixed effects for "years to BRC deposit"; the residuals from this regression are then grouped into quartiles to capture differences in the level of the predeposit impact of different root articles. In 6–3, we report the BRC deposit coefficient (ψ) for root articles in each quartile. The results suggest that the impact of BRC deposit is highest for articles from the "middle" of the quality distribution. In particular, there is a significant difference between the impact of BRC deposit on articles in the second and third quartiles, relative to the top quartile of predeposition citations. These findings are of particular interest to public policy: if the marginal impact of BRC deposit were concentrated exclusively among

[25] The Web of Science identifies each paper's "reprint author" as the individual to whom reprint copies of the paper are sent and to whom questions are addressed.

TABLE 7A—EXPLORING HETEROGENEITY IN TREATMENT EFFECTS BY CITING ARTICLES

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL [Incidence-rate ratios in brackets in top line] Estimated coefficients in second line (Block bootstrapped SEs reported in parentheses)			
	(7A-1)		(7A-2)	
	DV = Forward citations by articles not in top journals	DV = Forward citations by articles in top journals	DV = Forward citations by articles with a single subject field	DV = Forward citations by articles with multiple subject fields
Article characteristics				
BRC-ARTICLE,	[1.721]	[2.098]	[0.728]	[2.543]
POST-DEPOSIT	0.543	0.741	−0.317	0.933
	(0.193)***	(0.401)***	(0.137)*	(0.353)***
Control variables				
Age FEs	Sig.		Sig.	
Year FEs	Sig.		Sig.	
Observations	9,596		7,294	
Log likelihood	−14,891.16		−13,256.34	
Number of groups	426		323	
Test for equality of regression *BRC-ARTICLE, POST-DEPOSIT* coefficients				
	$\chi^2(1) = 0.70$		$\chi^2(1) = 32.91$	
	$\Pr > \chi^2 = 0.404$		$\Pr > \chi^2 = 0.000$	

Notes: Coefficients for BRC-window articles included in regressions but suppressed in order to focus on key variables in the analysis.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

articles with the highest level of predeposit citations, optimal policy might simply be to ensure accessibility and integrity for those discoveries that received the highest level of follow-on work after publication; however, since BRC deposit has a positive impact for all quartiles (and the highest marginal impact is associated with articles from the "middle" of the distribution), it may be important to ensure the accessibility of knowledge and materials even for discoveries that are not deemed to be particularly important in the period immediately after their publication.

Table 7 shifts the focus by examining whether BRC deposit has a differential level of impact on *different subpopulations of potential citers*. To evaluate heterogeneity among subpopulations, we first classify each citation to each article according to bibliometric characteristics, and then calculate the total number of citations received from that subpopulation in each year after publication. Specifically, we estimate

$$(4)\quad FORWARD\,CITATIONS_{i,l,t} = f(\varepsilon_{i,l,t}; \gamma_{i,l} + \beta_{t,l} + \delta_{t-pubyear,l} + \sum_{l=1,\ldots L} \psi_{Window,l}\, \mathbf{l}_l\, BRC\text{-}ARTICLE \times WINDOW\,PERIOD_{i,t} + \sum_{l=1,\ldots L} \psi_l\, \mathbf{l}_l\, BRC\text{-}ARTICLE \times POST\text{-}DEPOSIT_{i,t}),$$

where $\mathbf{l}_l$ is an indicator function equal to 1 for all sub-population citation years for sub-group l, ψ_l represents the impact of the treatment on sub-population l,

TABLE 7B—EXPLORING THE IMPACT OF DEPOSIT ON UNIQUE NEW CITATIONS

	CONDITIONAL FIXED EFFECTS NEG BINOMIAL [Incidence-rate ratios in brackets in top line] Estimated coefficients in second line (Block bootstrapped SEs reported in parentheses)		
	(7B-1) DV = annual count of unique new institutions in set of citing papers	(7B-2) DV = annual count of unique new journals in set of citing papers	(7B-3) DV = annual count of unique new countries in set of citing papers
Article characteristics			
BRC-ARTICLE, POST-DEPOSIT	[1.976] 0.681 (0.281)***	[1.737] 0.552 (0.223)***	[1.909] 0.647 (0.250)***
Control variables			
Age FEs	Sig.	Sig.	Sig.
Year FEs	Sig.	Sig.	Sig.
Observations	4,860	4,860	4,860
Log likelihood	−9,255.01	−7,305.66	−4,304.69
Number of groups	216	216	216

Notes: Coefficients for BRC-window articles included in regressions but suppressed in order to focus on key variables in the analysis. IRRs reported in brackets; raw coefficients reported in middle line.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

conditional on a fixed effect for citations by each subpopulation for each article, and allowing for separate article age and citation year fixed effects for each subpopulation. By evaluating how the impact of BRC deposit varies across different citation subpopulations (e.g., $H_0 : \psi_l = \psi_k$), we can evaluate whether the types of citations received as the result of BRC deposit seem to be linked with the role of BRCs in enhancing certification and enabling independent access to research materials.

For example, we calculate the number of citations to each article by articles in top-tier journals, and citations that are *not* in top-tier journals. We then specify a stacked regression, where the dependent variable in each group is the number of citations received from a given subpopulation in a given year. Each negative binomial regression includes separate fixed effects for each calendar year–subpopulation, article-age subpopulation, and conditional fixed effects for each root article–subpopulation. Consistent with the certification role of research-enhancing institutions, we are particularly interested in testing whether the increase in citations associated with BRC deposit are associated with higher-quality, more complex research projects.

We report results for two types of subpopulation groupings. In 7A-1, the citations from both top-tier journals and non–top-tier journals are estimated to increase after BRC deposit (the estimates imply a 110 percent citation boost from top-tier journals and a 72 percent citation boost from non–top-tier journal articles). Though the coefficient is larger for citations from top-tier journals, the difference between the coefficients is not statistically significant ($p = 0.40$).[26] In 7A-2, we investigate whether the

[26] This basic pattern of results obtains across a wide range of specifications that split citations by several different measures of perceived quality. For example, BRC deposit is estimated to result in a 98 percent increase in citations from reprint authors from top-50 universities compared with an 81 percent increase in citations by reprint authors outside of the top-50 universities. As in 7A-1, while each treatment coefficient is statistically significant, the coefficients are not statistically different from each other.

change in citations arising from BRC deposit is associated with simple versus complex research articles. We implement this distinction by examining citations either by articles that report a single subject field (e.g., biochemical research methods) or articles that report multiple subject fields (e.g., microscopy and parasitology). While single-subject field articles are more likely to be narrow papers on more modular topics, multisubject papers are more likely to be broader papers on more complex topics. The results here are striking. While multisubject articles are estimated to increase more than 150 percent after BRC deposit, single-subject articles are estimated to experience a modest decline (these coefficients are statistically different from each other). Taken together, Table 7A provides some additional evidence consistent with the certification hypothesis: the boost in citations resulting from BRC deposit is weakly associated with citations from high-quality follow-on research and is strongly associated with more complex research projects (where the reduction in uncertainty associated with using certified biological materials may have a higher marginal benefit). These results are also consistent with the hypothesis that the boost in citations in prior tables represents an expansion in follow-on research, rather than simple substitution in citations or materials. While substitution might result in higher citation counts for BRC-linked articles, the simplest forms of substitution should not significantly affect the character of follow-on research projects.

Finally, in Table 7B, we investigate whether the impact of BRC deposit is associated with an increase in the breadth of the research community building on a particular discovery. Similar to the substitution analysis in Table 5, our analysis is, in some sense, a falsification exercise, insofar as we can evaluate whether BRC deposit simply results in an increased number of citations, without really changing the portfolio of where those citations come from. Our approach is first to calculate, for each citation year, the number of citations that come from sources that had not referenced the root article in prior years. Specifically, for each citation year, we calculate the number of unique new institutions, unique new journals, and unique new countries. Each specification in Table 7B also includes article age, calendar year, and conditional article fixed effects. In each case, BRC accession corresponds to a statistically significant and quantitatively important expansion in the sources of citations for BRC-linked root articles. For example, 7B-1 suggests that BRC accession is associated with a 98 percent increase in the number of institutions citing BRC-linked root articles that had not previously cited those root articles. Though by no means dispositive, these findings are consistent with the idea that the independent access offered by BRCs to biological materials increases the exploitation of the knowledge associated with those materials by a broadened group of follow-on researchers.[27]

C. *Assessing the Cost-Effectiveness of BRCs*

Our final exercise is a "back-of-the-envelope" cost-effectiveness analysis, which is summarized in Table 8. A complete cost-benefit analysis is beyond the scope of our analysis, since we have no direct measure of the research productivity impact

[27] In a related setting (mouse genetics research), Murray et al. (2009) expand on this type of analysis to evaluate the impact of a shift in openness (resulting from a relaxation of intellectual property protection) on the restrictiveness of formal intellectual property rights on the diversity and novelty of upstream scientific research.

TABLE 8—BRC DEPOSIT COST-EFFECTIVENESS ANALYSIS

Calculation	Baseline cost per citation[a]	BRC accession cost	BRC citation boost	Cost per citation for BRC-linked articles	BRC cost-effectiveness index[c]
BRC-deposited articles citation boost	$2,887	$10,000	40.96	$244.12	11.83
Sample article citation boost	$2,887	$10,000	30.14	$331.80	8.70
Top ten university citation boost[b]	$2,887	$10,000	16.65	$600.45	4.81
Random university citation boost[b]	$2,887	$10,000	9.68	$1,032.94	2.79

Notes:

[a] Based on Adams-Griliches (1986) estimate of cost per citation.

[b] Based on Adams-Griliches (1986) estimate of citations received by articles authored by member of top ten biology departments and other university biology departments.

[c] BRC cost-effectiveness index = (baseline citation cost)/(BRC citation cost).

of BRCs. We are, however, able to undertake a simple comparison of the citation impact of expenditures targeted at accessioning biological materials into a BRC (i.e., ensuring that today's discoveries are accessible to follow-on researchers) versus funding for an additional research project. Specifically, we compare the "cost per citation" of funding new research studies on the one hand and accessioning biological materials associated with already published research on the other. Such a counterfactual is inherently speculative; thus, we choose benchmarks that reduce our estimate of relative cost-effectiveness of marginal investment in BRCs in comparison to marginal funding for additional research projects.

Our counterfactual requires an estimate of (i) the cost per citation induced by public research funding, (ii) the cost of BRC accession, and (iii) the number of citations induced by BRC accession. We set the cost per citation from research funding using the lowest estimate of this metric from James Adams and Griliches (1998) (corresponding to citations resulting from expenditures at a top ten biology department)—$2,400 in 1996 USD (which we adjust to $2,887 in 2002 USD using the BEA R&D price deflator). We set the cost of BRC accession to be equal to $10,000 per material (corresponding to the maximum of the range reported from survey evidence in OECD 2001). Finally, we draw on our estimate of the citation "boost" to compute the incremental number of citations expected to result from deposit and accession into a national BRC. Specifically, we use the estimate from 4–1 (67.7 percent, which includes a BRC-specific time trend and is lower than the baseline estimates in 3–4). We apply this treatment effect estimate to calculate the incremental citations arising from BRC deposit for four different "types" of research articles. Adams and Griliches (1998) offer two useful benchmarks for comparison: publications from a top ten biology department are associated with 24.6 citations during their first five years of publication, and publications from a biology department outside the top ten are associated with 14.3 citations during their first five years of publication. If we apply our treatment effect to these citation counts, BRC accession would be associated with 16.7 and 9.7 citations, respectively. Similarly, if we focus on all articles in our sample, the citation boost is estimated to be 30.1 (from a baseline of 60 citations over a five-year period), and the citation boost associated with

just the treatment articles is estimated to be 41.0 (from a baseline of 60 citations over a five-year period).

Dividing the BRC accession cost by the BRC citation boost yields an estimate of the BRC citation cost, which can be compared with the baseline citation cost. Across all counterfactuals, BRC accession is associated with a significant reduction in cost per citation. The estimates range from a factor of three (for a "random" article) to more than ten (for BRC-linked articles). While it is important to exercise caution in interpreting these estimates (citation impact is certainly not the only criterion for research investment productivity), it is useful to emphasize that a primary funding criterion of the NIH and related agencies is the potential for impact on future research (a criterion often assessed through simple citation counts). The analysis suggests that investments in research-enhancing institutions amplify the impact of research; the marginal NIH dollar may be more effectively spent on ensuring the accessibility and authenticity of research rather than simply funding additional research.

VI. Conclusion

In this paper, we propose a methodology to identify whether an institution exerts a positive externality on the accumulation of knowledge. We examine an institution that preserves, authenticates, and circulates life sciences research materials and find a substantial amplification in cumulative knowledge production. Our empirical approach combines large-scale citation analysis with a difference-in-differences approach to causal inference, allowing us to disentangle the impact of institutions on the dynamics of cumulative research, an approach that has since been adopted in an increasing number of papers (Murray and Stern 2007; Ajay Agrawal and Avi Goldfarb 2008; Rysman and Simcoe 2008; Murray et al. 2009; Furman, Murray, and Stern 2010). Over the past several years, science funding agencies (including the National Science Foundation and National Institutes of Health) have placed significant priority on the development of the "Science of Science Policy" (SOSP) (John H. Marburger III 2005; Jaffe 2006), in which the tools of program evaluation can be used to evaluate alternative institutional arrangements and science policy choices. One contribution of this paper is the introduction of the combination of citation analysis with a difference-in-differences approach as a SOSP methodology.

Our findings bear directly on public policy toward the preservation, certification, and distribution of biological materials, data, and resources. The policy issues concerning biological materials and data cut across a wide range of policy areas, from federal funding for embryonic stem cell research (where the lack of federal funding for new cell lines may have affected the rate of scientific progress (Furman, Murray, and Stern 2010)), to public investment in freely accessible databases such as the Human Genome Project (Kyle Jensen and Murray 2005), to the potential for conflict between national security and academic freedom in bioweapons research, illustrated most directly in the case of the identification of anthrax strains associated with the 2001 attacks (Stern 2004). In each of these cases, there is a significant gap between the public and private incentives to make authenticated biological materials and data available on an independent basis to follow-on researchers (Mukherjee and Stern 2009; Carolin Häussler 2010). Our findings offer support for policy proposals that (i) premise public funding or publication of research on a commitment to provide access to that knowledge to

future scientific researchers (Nardone 2007) and (ii) shift funding priorities on the margin away from simply funding research projects to funding research streams that accumulate a body of systematized knowledge that is available on an open-access basis. More generally, the analysis highlights the crucial role of openness and independent access as prerequisites for cumulative knowledge production and suggests the value of research identifying the economic conditions and empirical circumstances that allow Open Science to succeed as an economic institution (Mokyr 2002; Aghion, Dewatripont, and Stein 2008; David 2001; Murray et al. 2009; Heidi Williams 2010).

REFERENCES

Adams, James D., and Zvi Griliches. 1998. "Research Productivity in a System of Universities." *Annales d'Economie et de Statistique,* (49–50): 127–62.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica*, 60(2): 323–51.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic Freedom, Private-Sector Focus, and the Process of Innovation." *RAND Journal of Economics*, 39(3): 617–35.

Agrawal, Ajay, and Avi Goldfarb. 2008. "Restructuring Research: Communication Costs and the Democratization of University Innovation." *American Economic Review*, 98(4): 1578–90.

Allison, Paul D., and Richard P. Waterman. 2002. "Fixed-Effects Negative Binomial Regression Models." In *Sociological Methodology,* Vol. 32, ed. Ross M. Stolzenberg, 247–65. Boston: Blackwell Publishing.

American Type Culture Collection. 2009. "Who We Are." www.atc.org. (accessed March 29, 2009).

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–75.

Branstetter, Lee G., and Mariko Sakakibara. 2002. "When Do Research Consortia Work Well and Why? Evidence from Japanese Panel Data." *American Economic Review*, 92(1): 143–59.

Bush, Vannevar. 1945. *Science: The Endless Frontier.* Washington, DC: United States Government Printing Office.

Campbell, Eric G., Brian R. Clarridge, Manjusha Gokhale, Lauren Birenbaum, Stephen Hilgartner, Neil A. Holtzman, and David Blumenthal. 2002. "Data Withholding in Academic Genetics: Evidence from a National Survey." *Journal of the American Medical Association,* 287(4): 473–80.

Chatterjee, Rhitu. 2007. "Cell Biology: Cases of Mistaken Identity." *Science,* 315(5814): 928–31.

Cockburn, Iain M., and Rebecca M. Henderson. 1998. "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery." *Journal of Industrial Economics*, 46(2): 157–82.

Cohen, Wesley M., and John P. Walsh. 2008. "Real Impediments to Academic Biomedical Research." In *Innovation Policy and the Economy,* Vol. 8, ed. Adam B. Jaffe, Josh Lerner, and Scott Stern, 1–30. Chicago: University of Chicago Press.

Cypress, Raymond. 2003. "Biological Research Centers: From Concept to Reality." Manassas, VA: ATCC.

Dasgupta, Partha, and Paul A. David. 1994. "Toward a New Economics of Science." *Research Policy,* 23(5): 487–521.

David, Paul A. 2001. "From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science.'" Stanford University Economics Working Paper 01-006.

de Solla Price, Derek. 1976. "A General Theory of Bibliometric and Other Cumulative Advantage Processes." *Journal of the American Society for Information Science,* 27(5): 292–306.

Evans, James. 2004. "Sharing the Harvest: The Uncertain Fruits of Public/Private Collaboration in Plant Biotechnology." Unpublished.

Furman, Jeffrey L., and Scott Stern. 2006. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." National Bureau of Economic Research Working Paper 12523.

Furman, Jeffrey L., and Scott Stern. 2011. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research: Dataset." *American Economic Review,* http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.1933.

Furman, Jeffrey L., Fiona E. Murray, and Scott Stern. 2010. "Growing Stem Cells: The Impact of US Policy on the Geography and Organization of Scientific Discovery." Unpublished.

Gallini, Nancy, and Suzanne Scotchmer. 2002. "Intellectual Property: When Is It the Best Incentive System?" In *Innovation Policy and the Economy,* Vol. 2, ed. Adam B. Jaffe, Josh Lerner, and Scott Stern, 51–77. Cambridge, MA: MIT Press.

Gans, Joshua S., and Scott Stern. 2003. "The Product Market and the Market for 'Ideas': Commercialization Strategies for Technology Entrepreneurs." *Research Policy,* 32(2): 333–50.

Garfield, Eugene. 1979. "Is Citation Analysis a Legitimate Evaluation Tool?" *Scientometrics,* 1(4): 359–75.
Garfield, Eugene, and Uriel H. Schoenbach. 1956. "Citation Indexes for Science." *Science,* 123(3185): 61–62.
Gold, Michael. 1986. *A Conspiracy of Cells.* Albany, NY: SUNY Press.
Greene, William. 2004. "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model." *Econometric Reviews*, 23(2): 125–47.
Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28(4): 1661–707.
Griliches, Zvi. 1998. *R&D and Productivity: The Econometric Evidence*. Chicago: University of Chicago Press.
Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.
Hall, Bronwyn H., Jacques Mairesse, and Laure Turner. 2007. "Identifying Age, Cohort, and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists." *Economics of Innovation and New Technology*, 16(1–2): 159–77.
Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, 52(4): 909–38.
Häussler, Carolin. 2010. "The Determinants of Information Sharing: An Analysis of Firm and University Scientists." Unpublished.
Heller, Michael A., and Rebecca S. Eisenberg. 1998. "Can Patents Deter Innovation? The Anticommons in Biomedical Research." *Science,* 280(5364): 698–701.
Henderson, Rebecca, Adam B. Jaffe, and Manuel Trajtenberg. 1998. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988." *Review of Economics and Statistics*, 80(1): 119–27.
Hunter, Cevera, Jenny C. 1996. "The Importance of Culture Collections in Industrial Microbiology and Biotechnology." In *Culture Collections to Improve the Quality of Life,* ed. R. A. Sampson, J. A. Stalpers, D. van der Mei et al, 158–62. Baarn: Centraalbureau voor Schimmelcultures.
Inwood, Stephen. 2003. *The Forgotten Genius: The Biography of Robert Hooke 1635–1703.* San Francisco: MacAdam/Cage Publishing.
Irwin, Douglas A., and Peter J. Klenow. 1996. "High-Tech R&D Subsidies: Estimating the Effects of Sematech." *Journal of International Economics*, 40(3–4): 323–44.
Jaffe, Adam B. 2006. "The Science of Science Policy." Keynote Address, NSF Workshop on Advancing Measures of Innovation: Knowledge Flows, Business Metrics, and Measurement Strategies, June 6, 2006, Arlington, VA.
Jaffe, Adam B., and Josh Lerner. 2001. "Reinventing Public R&D: Patent Policy and the Commercialization of National Laboratory Technologies." *RAND Journal of Economics*, 32(1): 167–98.
Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, 108(3): 577–98.
Jensen, Kyle, and Fiona Murray. 2005. "Intellectual Property Landscape of the Human Genome." *Science,* 310(5746): 239–40.
Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy*, 103(4): 759–84.
Kortum, Samuel, and Josh Lerner. 2000. "Assessing the Contribution of Venture Capital to Innovation." *RAND Journal of Economics*, 31(4): 674–92.
Lamoreaux, Naomi R., and Kenneth L. Sokoloff. 1999. "Inventive Activity and the Market for Technology in the United States, 1840–1920." National Bureau of Economic Research Working Paper 7107.
Lerner, Josh, and Jean Tirole. 2004. "Efficient Patent Pools." *American Economic Review*, 94(3): 691–711.
Levin, Sharon G., and Paula E. Stephan. 1991. "Research Productivity over the Life Cycle: Evidence for Academic Scientists." *American Economic Review*, 81(1): 114–32.
Leydesdorff, Loet. 2001. *The Challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications.* Parkland, FL: Universal Publishers.
MacKinnon, James G. 2002. "Bootstrap Inference in Econometrics." *Canadian Journal of Economics*, 35(4): 615–45.
MacLeod, Roderick A. F., Wilhelm G. Dirks, Yoshinobu Matsuo, Maren Kaufmann, Herbert Milch, and Hans G. Drexler. 1999. "Widespread Intraspecies Cross-Contamination of Human Tumor Cell Lines Arising at Source." *International Journal of Cancer,* 83(4): 555–63.
Marburger, John H., III. 2005. "Wanted: Better Benchmarks." *Science,* 308(5725): 1087.
Masters, John R. 2002. "Timeline: HeLa Cells 50 Years on: The Good, the Bad and the Ugly." *Nature Reviews Cancer,* 2(4): 315–19.
Mehta, Aditi, Marc Rysman, and Tim Simcoe. 2010. "Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion." *Journal of Applied Econometrics*, 25(7): 1179–204.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations,* ed. Norman Storer. Chicago: University of Chicago Press.
Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy.* Princeton, NJ: Princeton University Press.
Mowery, David C., and Arvids A. Ziedonis. 2007. "Academic Patents and Materials Transfer Agreements: Substitutes or Complements?" *Journal of Technology Transfer,* 32(3): 157–72.
Mowery, David C., Richard R. Nelson, Bhaven N. Sampat, and Arvids 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(1): 99–119.
Mukherjee, Arijit, and Scott Stern. 2009. "Disclosure or Secrecy? The Dynamics of Open Science." *International Journal of Industrial Organization,* 27(3): 449–62.
Murray, Fiona. 2010. "The Oncomouse that Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions." *American Journal of Sociology,* 116(2): 341–88.
Murray, Fiona, and Scott Stern. 2007. "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis." *Journal of Economic Behavior and Organization*, 63(4): 648–87.
Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern. 2009. "Of Mice and Academics: The Role of Openness in Science". Unpublished.
Nardone, Roland M. 2007. "Eradication of Cross-Contaminated Cell Lines: A Call for Action." *Cell Biology and Toxicology,* 23(6): 367–72.
Nelson, R. R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy*, 67(6): 297–306.
Nelson, Richard R. ed. 1993. *National Innovation Systems: A Comparative Analysis*. New York: Oxford University Press.
OECD. 2001. *Biological Research Centers: Underpinning the Future of Life, Sciences and Biotechnology.* Paris: OECD.
Pavitt, Keith, and Pari Patel. 1988. "The International Distribution and Determinants of Technological Activities." *Oxford Review of Economic Policy*, 4(4): 35–55.
Powell, Walter W. 1998. "Learning from Collaboration: Knowledge and Networks in the Biotechnology and Pharmaceutical Industries." *California Management Review,* 40(3): 228–40.
Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71–102.
Rosenberg, N. 1963. "Technological Change in the Machine Tool Industry, 1840–1910." *Journal of Economic History*, 23: 414–46.
Rosenberg, Nathan. 1979. "Technological Interdependence in the American Economy." *Technology and Culture,* 20(1): 25–50.
Rosenkopf, Lori, and Michael L. Tushman. 1998. "The Coevolution of Community Networks and Technology: Lessons from the Flight Simulation Industry." *Industrial and Corporate Change,* 7(2): 311–46.
Rysman, Marc, and Timothy Simcoe. 2008. "Patents and the Performance of Voluntary Standard-Setting Organizations." *Management Science,* 54(11): 1920–34.
Sakakibara, Mariko, and Lee 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, Suzanne. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives*, 5(1): 29–41.
Simkin, M. V., and V. P. Roychowdhury. 2003. "Read before You Cite!" *Complex Systems,* 14(3): 269–74.
Skloot, Rebecca. 2010. *The Immortal Life of Henrietta Lacks.* New York: Crown Publishing.
Smith, David. 2003. "Culture Collections over the World." *International Microbiology,* 6(2): 95–100.
Sorenson, Olav, and Lee Fleming. 2004. "Science and the Diffusion of Knowledge." *Research Policy,* 33(10): 1615–34.
Stern, Scott. 2004. *Biological Resource Centers: Knowledge Hubs for the Life Sciences.* Washington, DC: Brookings Institution Press.
Walsh, John P., Charlene Cho, and Wesley M. Cohen. 2005. "View from the Bench: Patents and Material Transfers." *Science,* 309(5743): 2002–03.
Williams, Heidi. 2010. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." Unpublished.
Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
World Federation of Culture Collections, World Data Center for Microorganisms. "The Culture Collection in this World." WDCM Statistics, http://wdcm.nig.ac.jp/statistics.html (accessed March 29, 2009).
Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." American Economic Review, 88(1): 290–306.