

**A PENNY FOR YOUR QUOTES?  
THE IMPACT OF BIOLOGICAL RESOURCE CENTERS  
ON LIFE SCIENCES RESEARCH**

**Jeffrey L. Furman  
Boston University**

**Scott Stern  
Northwestern University and NBER**

**March 2004**

***Chapter 4 in  
Biological Resource Centers: Knowledge Hubs for the Life Sciences***

**Scott Stern**

**(The Brookings Institution: Washington, DC, *forthcoming* 2004)**

## A. Introduction

The qualitative analysis suggests that knowledge and materials associated with biological resource centers have an important impact on the process of cumulative research that is so central to scientific advance, technological progress, and long-term economic prosperity. The framework in Chapter 3 identifies four distinct advantages associated with BRCs: authentication and certification of deposits, the long-term preservation of materials, independent access, and the exploitation of scale and scope economies. As a cumulative knowledge hub, BRCs amplify the eventual impact of a given piece of research, over a longer period of time and across a wider swath of the scientific community.

This chapter evaluates these claims through a quantitative investigation of the impact of BRCs on the cumulateness of research in the life sciences. Our approach exploits two insights about the nature of life sciences research:

- citations (either to scientific publications or patents) can serve as a useful though imperfect measure of the cumulative impact of discoveries
- specific BRC deposits can be linked to scientific publications or patents (and the timing of some BRC deposits is distinct from the date of this publication )

By undertaking a citation analysis of publications associated with BRC deposit, this analysis exploits the fact that the process of knowledge accumulation in the life sciences is made “visible” through the pattern of article citations in the scientific literature. Indeed, over the past decade, researchers in economics and other social sciences have increasingly exploited patterns of citation and referencing to gain insight into how scientific and technical knowledge is created and distributed (Griliches, 1994). Perhaps the key insight of this economic and sociological research is that the number of times an article is cited (and by whom and how long after its

publication) provides a useful (though noisy) index of the “impact” of that article on subsequent scientific research (See Box 4A, “The Use (and Misuse) of Citation Analysis”).

We extend these prior analyses by linking scientific articles (and the knowledge associated with those articles) to specific BRC deposits. The analysis takes advantage of the fact that the leading biological resource center in the United States, the ATCC, collects bibliographical reference information for the majority of deposits in its collection (See Box 4B, “Choosing the Sample” and Appendix 4A). In other words, the ATCC catalogue includes the name of the original depositor, the date of the deposit, and key scientific articles associated with materials in its collections. Moreover, in some cases, the date of the deposit into the ATCC is distinct from the date of initial scientific publication, offering a method for isolating the specific impact of BRC deposit. We exploit each of these features to develop a number of datasets and assess a number of key hypotheses about the impact of BRCs on cumulative research (See Box 4C, “Citation Analysis Data and Measures”).

In particular, if depositing biological materials in BRCs increases the accessibility of that knowledge to future research (i.e., increases its cumulateness), then scientific articles associated with BRC deposits should be more intensively cited as a result of their greater impact on follow-on research. Since articles that are associated with BRC deposits may simply be more influential pieces of scientific research, it is critical to identify a “control” group of articles of similar scientific “quality” and initial publication date that are not associated with a BRC deposit. As well, for policy analysis, it is important to disentangle the intrinsic scientific importance of articles from the specific impact of BRC deposit. If research is more likely to be cited (and over a longer period of time) when associated materials are deposited in a BRC, then BRCs exert an impact on the cumulateness of life sciences research.

Our empirical assessment is divided into three parts. First, we evaluate the citations of research publications linked to BRC deposits, relative to publications from the same journals published at the same time frame. The results are dramatic: BRC-linked articles have more than a 200% greater likelihood of being cited, a gap that becomes larger with the length of time since initial publication. This citation advantage to BRC-linked articles is evident across several different collections of the ATCC, and has increased during the 1990s. The lead authors of BRC-linked articles tend to be located in the United States (not surprisingly, given that the sample is drawn from ATCC deposits) but are otherwise similar to lead authors of the control articles. A smaller, more preliminary study of patents that draw on BRC materials reinforces the increasing role that BRCs are playing in biotechnology innovation.

The second stage of the analysis focuses on estimating the “marginal” impact of BRCs in explaining the large citation advantage to BRC-linked articles. By comparing how articles’ citation patterns change *after* materials associated with those articles are deposited in a BRC, the specific role played by BRCs in amplifying the impact of research is isolated.<sup>1</sup> The results indicate that articles associated with BRC materials experience a greater than 80% boost in their citations when those materials are deposited in ATCC. Moreover, BRC deposit increases the “longevity” of research impact – BRC-lined continue to receive citations at a substantially higher rate, even twenty years after materials have been deposited. The “boost” from BRC deposit has also become far more pronounced since the mid 1980s. BRC deposit affects not only the number of citations change but also the type of citations; citations to articles after materials are deposited in a BRC tend to have higher citation rates themselves. Overall, as an economic institution, BRCs play an important and unique role in promoting step-by-step progress in the life sciences.

---

<sup>1</sup> As discussed in some more detail below, our methodology is to estimate a “differences-in-differences” estimator, taking advantage of features of how deposits are made for specific collections.

Finally, a “rate of return” analysis is presented, assessing how expenditures on BRC deposit compare to alternative research investments in terms of promoting cumulative progress. Expenditures aimed at increasing the impact of published research are compared with expenditures aimed at encouraging new publications. The results are dramatic. While the average “cost” per citation for university-based biological research is estimated at \$2400, the cost per citation induced by BRC deposit is less than \$900 (under the most expensive calculation). Relative to traditional grant activities, investments in BRC authentication, storage and access activities dramatically improve the accessibility of knowledge and so facilitate cumulative knowledge production. This core finding motivates the detailed policy analysis conducted in Chapter 5.

## **B. The Citation Patterns of BRC-Linked Articles**

Our initial analysis compares the rate of citation to articles linked to ATCC deposits with groups of control articles judged to be of similar scientific quality. We construct article “sets,” matching each BRC-linked article with a “nearest neighbor” article published in the same issue of the same journal as well as a “most related” article published in the same volume of the same journal. Because both BRC-affiliated article and control articles have undergone the same scientific review process (and been judged to contain roughly similar scientific merit), patterns of citations by future researchers to these articles provide an indication of their relative impact on subsequent research. As well, because each pair comes from a single issue (or the same volume) of a journal, the “opportunity” to be cited by future researchers is virtually identical within each set.<sup>2</sup> Several striking findings emerge:

---

<sup>2</sup> BRC-linked articles may be published a few months earlier or later in the volume than “most-related” article controls. Unless there is some reason to believe that BRC-linked articles are published earlier in the year than

- *BRC-linked articles have a much higher rate of citation than articles published in the same volume of the same scientific journal.* While both the BRC-linked articles and the control samples have comparable numbers of Backward Citations (i.e., they cite similar number of references), BRC-linked articles have a Forward Citation rate nearly triple that of control articles published on the pages that precede them (*nearest neighbor article controls*) and nearly double that of the most similar articles that appear in the same volume of the same journal (*most-related article controls*) (Figure 4A). By 2001, BRC-linked articles receive nearly 90 more cumulative Forward Citations on average than either type of control article (Figure 4B).
- *This citation “gap” appears across article samples linked to several distinct ATCC collections.* Figure 4C divides up the article pairs according to whether articles are linked to deposits in the Cell Biology, Bacteriology, and Molecular Biology collections of the ATCC and compares citations for BRC-linked articles with those of the *nearest neighbor controls*. The advantage experienced by ATCC-linked articles appears for all collections, with the Cell Biology collection articles experiencing the highest overall boost.
- *The divergence in citations between BRC-linked articles and the control samples grows with the time since publication.* Perhaps no comparison exemplifies better how ATCC-linked articles differ from control articles than Figure 4D, which charts articles’ forward citation rates as their dates of publication become more distant. In the first few years after publication, BRC-linked article receive substantially more citations than those of control articles. As time passes, all sets of articles receive

---

“most-related” control articles, this would not bias the results. (Note that there appears to be no reason to believe that BRC-linked articles are published earlier or later in the year than “most related” control articles.)

fewer citations; however, the gap in citations between those received by BRC-linked and control articles increases substantially – reaching at least 250% ten years after publication. Even twenty years after publication, BRC-linked articles continue to average significant numbers of citations, while control articles receive only one or two. In other words, although each set of matched articles is judged at the time of publication to hold similar scientific merit (i.e., they cleared the same publication “hurdle”), the knowledge described in BRC-linked articles tends to exert a more pervasive impact on the research process for a much longer length of time.

- *ATCC-linked articles have similar “characteristics” to the control articles except that lead authors are more likely to be located in the United States. Even after controlling for these differences in author and article characteristics, the citation gap between BRC-linked and control articles remains high.* As part of our examination of this data, we have undertaken a number of tests to investigate how BRC-linked articles differ from our control sample. Overall, the two samples seem similar, with the exception that control articles tend to be published by non-US authors, and that one control set averages somewhat fewer pages and more co-authors (Figure 4E). Even after accounting for detailed characteristics such as these in our analysis, the citation advantage experienced by BRC-linked articles remains similarly high.
- *An increasing number of patents acknowledge BRCs, such as ATCC, in their scientific citations, abstracts, and descriptions. These BRC-referencing patents appear similar to other patents in terms of commercial and technological significance, although they do make more references to academic articles.* To investigate the role that BRCs seem to have increasingly played in biotechnology, we

also have assessed whether ATCC reference materials contribute to the development of patented life sciences technologies. In particular, we track citations to ATCC materials in four key patent classes where biotechnology innovation is most prevalent (Figure 4F). The results demonstrate a dramatic increase in the fraction of patents that reference ATCC materials between 1981-2002. Taking a more in-depth look at 80 of these patents (drawn evenly from the four classes), we observe that patents that acknowledge their use of ATCC materials are more likely to reference non-patent materials (like academic articles) in their applications. Except for the fact that they make more intensive use of science to develop technology, these patents make references to prior patents and receive references from future patents in numbers similar to a control group of patents issued in the same technology classes at similar times (Figure 4G).

### **C. The Impact of BRC Deposit on Scientific Accessibility**

The analysis points to a dramatic difference in the forward citation patterns of articles linked to BRC deposits, even compared to articles judged to pass the same publication hurdle. However, the approach so far does not isolate out the “marginal” impact of BRCs in fostering cumulative knowledge production. Specifically, it is possible that BRC-linked articles are of higher scientific importance than the “control” articles we compared them with, notwithstanding their publication in the same issue of the same journal. To disentangle these two different effects that may both lead BRC-linked articles to be more highly cited, we turn to a nuanced, but

ultimately more revealing, empirical methodology. This methodology measures how the pattern of citation to a scientific article *changes* after the materials associated with that article are deposited in a BRC. By focusing on the changing citation pattern that results from deposit, we can infer the direct impact that BRCs have on the accessibility of knowledge.

The approach takes advantage of the fact that, over the past twenty years, a number of “orphan” collections have been “adopted” by the large national BRCs. More specifically, collections maintained in a specific university laboratory or research institute have suffered from a funding crisis (or the retirement of key personnel), necessitating a search for a new home (See Box 4D, “The Special Collections”). In other words, these materials (often more than 100 distinct cultures) have been transferred from the peer-to-peer network to a knowledge hub, often years after the initial publication of articles associated with their identification and characterization. This set of circumstances – where materials are located for a period of time within the peer-to-peer network and then become accessible through a knowledge hub – allows us to assess the specific impact that deposit in a BRC has on the accessibility of knowledge (See Box 4E, “The Research Methodology,” Appendix 4A, and Furman and Stern (2003). Several findings from this analysis stand out:

- *Depositing materials in a BRC results in a substantial “boost” to the citations received by articles associated with that deposit.* As described in the accompanying box (“The Research Methodology”), our methodology exploits “special collections” whose materials are linked to articles where the timing of initial publication is earlier than the timing of BRC deposit. To estimate the “marginal” impact of BRC deposit on yearly citations, a regression procedure is employed that accounts for the amount of time since the article’s initial publication, the years in which citations occur, and

the overall impact of the article (independent of whether associated materials have yet been deposited in a BRC). This citation analysis procedure measures two key values:

- the degree to which BRC-linked articles simply represent more important scientific research (“Selection Effect”)
- the degree to which deposit of materials in the BRC increase the use of the knowledge produced in those articles by members of the scientific community (“Impact of BRC Deposit”)

These two drivers of the citation advantage of BRC-linked articles are displayed in Figure 4H. Most importantly, while the selection effect is large (accounting for an over 120% boost in citations), the impact of BRC deposit is measured to result in an additional 102% citation increase. In other words, when materials are deposited in a BRC, articles associated with those materials experience a doubling of their citation rate (relative to the pattern that would have been expected in the absence of BRC deposit).

- *The divergence in citations resulting from BRC deposit grows over time after the deposit occurs.* The overall advantage resulting from BRC deposit is not an immediate effect but a gradual growth that lasts over a decade. Figure 4I illustrates that in the decade prior to deposit, articles associated with the special collections do not experience an elevated rate of citation. However, within four years after deposit, these articles experience nearly a 50% boost, growing to approximately 120% within ten years after deposit. Consistent with the key role that preservation and storage play in an effective knowledge hub, BRCs amplify the impact of scientific research findings over the long term.

- *The citation gap associated with BRC deposit has increased markedly since the mid 1980s.* We earlier observed that BRC-linked articles have increasingly experienced a gain in their citation rates, starting during the mid-1980s. Figure 4J demonstrates this finding, focusing on how the “marginal” impact of BRCs has changed over time. The overall trend in the impact of ATCC deposit has been upward over time, and the last decade has witnessed a stronger impact than during the 1980s. This recent increase in impact highlights the central importance of certified biological materials in an era in which bioinformatics and genomics play an increasingly important role in life sciences research.
- *The boost from inclusion in a BRC is apparent across a number of distinct special collections.* The examination here draws from three distinct “orphan” collections transferred to the ATCC. Each of the collections that has been housed at ATCC for more than a decade experiences a substantial advantage in the citation rates of associated articles as a result of BRC deposit (Figure 4K). While the Gazdar collection has only been housed within ATCC for a few years (and so it may not be surprising that it has yet to experience a substantial overall boost), the HTB and TIB collections obtain citation boosts of approximately 56% and 105%, respectively.
- *For articles associated with BRC deposits, citations tend to come from articles that are themselves highly cited.* Not only do BRC-linked articles obtain higher rates of citation than control articles, but so too do the “second generation” articles that cite BRC-linked research (Figure 4L). This result is consistent with a view of BRCs as knowledge hubs that promote the diffusion of leading-edge scientific research. An analysis of citation patterns shows that those articles that cite BRC-linked articles are

otherwise similar to articles that cite *nearest neighbor* controls, with the exception that a somewhat higher fraction of articles that cite BRC-linked research come from US authors and that these articles have later average publication dates. The latter fact suggests that BRC-associated research streams enjoy greater longevity.

#### **D. The Cost-Effectiveness of BRCs as Cumulative Knowledge Hubs**

The results so far provide the first-ever quantitative evidence of the role that BRCs play as knowledge hubs in the life sciences. Articles linked to materials deposited at the ATCC have a much higher rate of citation and this citation boost can be specifically attributed to BRC deposit (rather than simply being associated with intrinsically more important science).

However, these findings cannot offer effective policy guidance without first at least a rough assessment of whether the benefits of BRC deposit are worth the costs that must be incurred to realize them. In the current context, a comprehensive cost-benefit analysis is beyond the scope of our analysis, since we cannot fully capture the degree to which access to BRC materials improves research productivity of users. However, it is possible to calculate a “back-of-the-envelope” estimate of the cost-effectiveness of BRC deposit, relative to public investments in traditional grant activities. As well, though revealing about the impact of BRC deposits at current collection levels, the benefits of BRC deposit might ultimately run into diminishing returns; this counterfactual analysis is therefore most revealing about the rate-of-return of incremental expenditures but may not capture the dynamics of large-scale shifts in the size and scope of BRC collections.

The knowledge hub framework highlights the fact that the ultimate social value of publicly funded research is closely linked to the exploitation of that research in the efforts of

future researchers. Moreover, within a particular discipline, citations to an article offer a noisy but useful proxy of the degree to which this cumulative process is occurring. While caution is important in interpreting citation-based results, the framework suggests that systematic increases in the rate of citation for a given type of paper is evidence for a higher rate of exploitation of that knowledge by the research community.<sup>3</sup>

This insight is at the heart of our cost-benefit calculation. Specifically, we compare how a given level of expenditures on BRC deposit compare to alternative research investments in terms of promoting cumulative progress. In other words, how do investments in BRC deposit and authentication activities compare to traditional grant programs in terms of their efficiency in seeding the knowledge stock of future researchers?

This exercise involves the calculation of three estimates:

- **Baseline Citation Cost:** The “cost per citation” paid by public funding agencies (such as NIH) when allocating resources that result in published scientific articles. This estimate is calculated using the estimates in a recent paper by Adams and Griliches (1996). Using data drawn from the 1980s, these authors estimate the relationship between expenditures and academic research output (papers and citations) for individual academic departments at top universities across the United States, including biology departments. Using these measures (and converting all expenditures into 1987 current dollars), they estimate the cost per citation to be \$2400 for expenditures at a top-ten biology department and at \$4200 for citations at non-elite public universities. Using the BEA R&D price deflator to recompute this number in current dollars, the *lowest* Adams and Griliches estimates of current cost

---

<sup>3</sup> As emphasized earlier, we are not discussing the citation rate of an individual paper, but the average citations rate associated with classes of papers distinguished according to specific criteria and sampling.

per citation is \$2887. Being conservative (in terms of estimating the effectiveness of BRC expenditures), we choose the lowest estimated cost per citation among these figures, and so set the Baseline Citation Cost at \$2400 for the life sciences.

- **BRC Accession Cost:** The full cost of deposit and accession into a national BRC collection such as the ATCC. The recent OECD Report on Biological Resource Centers provides estimates of this cost based on a recent survey; the highest estimate of BRC Accession Cost according to the OECD report is \$10,000 (this was the maximum of the range of the survey response given by the ATCC) (OECD, 2001). While it is likely that the true marginal accession cost may be somewhat lower than \$10,000, we use this high-end figure to bias us away from finding evidence for cost-effectiveness on the part of BRCs.
- **BRC Citation Boost:** The incremental number of citations expected to result from deposit and accession into a national BRC. We compute three different estimates of the BRC Citation Boost. The first two of these computations builds on the data provided by Adams and Griliches (1996). In their work, the average biology publication received 24.6 citations during the first five years of publication if authors were located at a top ten university and 14.3 citation if authors were located at universities below the top ten (in biology). As well, in our most conservative estimate, BRC deposit was associated with an 82% increase in citations. If we assume that the marginal accessioned material comes from a top ten university laboratory, then the marginal impact from deposit is estimated to be 20.2; if the accessioned materials is drawn truly at random, we assign the citation impact to be 11.7, based on the citation rates of articles published by authors outside the top ten.

We also compute the BRC Citation Boost directly from the estimates provided in the last section, focusing on the incremental boost realized by BRC-linked articles within the sample. Using this formulation, the BRC Citation Boost is 11.1; interestingly, BRC-linked articles within the sample have a BRC Citation Boost quite close to the estimated BRC Citation Boost for articles which would be drawn from random biology departments.

Dividing the BRC Citation Boost by the BRC Accession Cost yields an estimate of the BRC Citation Cost, which we can then compare with the Baseline Citation Cost. The results of this calculation are presented in Figure 4M. These estimates are dramatic. Even imposing the estimates that result in a conservative calculation, BRC deposit expenditures offer a minimum of a 267% efficiency benefit in terms of inducing citations. Of course, these calculations must be interpreted cautiously given the noisiness of citation data. However, to the extent that the primary criterion for current public basic research expenditures at NIH is the likelihood that such research will have important disciplinary impact (and that such an impact is often measured through citation counts), this analysis suggests that the impact of already funded and published research may be amplified cost-effectively as the result of the authentication, storage, and independent access associated with centralized cumulative knowledge hubs such as national biological resource centers. It is this core finding that motivates the detailed policy analysis in the next chapter.

## CHAPTER 4

### BOX 4A

#### THE USE (AND MISUSE) OF SCIENTIFIC PAPER AND PATENT CITATIONS

The use of citation analysis to evaluate the impact of scientific research and technological innovation goes back to the pioneering work of Eugene Garfield in the 1950s (who himself proposed and developed the *Science Citation Index*), and has experienced an explosion over the last two decades. Starting in the early 1980s, economists, sociologists, and policy analysts have increasingly exploited rich bibliometric datasets and sophisticated empirical toolkits to compile citation patterns that highlight key phenomena relating to scientific research and technological innovation. For example, for each scientific discipline, there exists a “hierarchy” of journals to which research papers may be submitted (and subjected to a refereeing process). Sociologists of science have documented the self-reinforcing nature of this publication hierarchy by showing that more prestigious journals tend to be associated with a higher rate of citation, which encourages a cycle of high-quality submissions. Economists, on the other hand, have exploited scientific and patent citation data in order to assess the importance of localized knowledge spillovers, intellectual property, and the impact of laws such as the Bayh-Dole Act (Griliches, 1984; Jaffe, 1986; Pakes, 1986; Jaffe, Henderson and Trajtenberg, 1993). Similarly, building on the pioneering database and analytical work of firms such as CHI Research, the National Science Foundation has increasingly used citation analysis for assessment purposes; for example, in order to measure the relative scientific capabilities of nations, the NSF highlights the role of citations as “an indicator of the perceived influence of a nation’s scientific outputs...and technical work.” (National Science Foundation, 2002).

Alongside this explosion in the use of citation data has come increased understanding of appropriate (and not so appropriate) uses of these statistics. Three methodological issues are central. First, small differences in the citation rate of a single paper (particularly early in its publication history) are of limited value in distinguishing the importance of research or its use by the research community. Simply put, the publication process is subject to delay and errors in attribution, and so citation data will tend to be noisy for a single or small group of papers. Appropriate use of citation measures are much more ideally suited to compare the overall impact of two groups of papers with similar *ex ante* characteristics. For example, in our present analysis, we compare the citations to papers that have been published in the same issue of the same journal but differ by whether the materials in the articles have been deposited within a BRC. Second, it is important to distinguish the interpretation of paper and patents citations. While paper citations usually recognize the key perceived influences on work by the author (but have no direct impact on the value or credit given the work), patent citations have a legal status that “bound” the new invention (since one cannot receive rights for intellectual property which has already been granted to others). This analysis focuses mostly on citations to scientific research papers, examining how the rate of citation differs according to whether materials described in articles are deposited in a biological resource center. Finally, nearly all citation databases are highly skewed; a very small number of articles (or patents) receive a very high share of the citations. This fact suggests that empirical findings should be checked to ensure they are not being driven by a small number of outliers, and that empirical procedures should be chosen that are appropriate for such skewed data. As we discuss in Box 4E and Appendix 4A, our use of “negative binomial count data” models allows us to account for these features of paper citation data.

Overall, the approach pursued here is premised on the idea that paper and patent citations provide a useful though quite noisy estimate of the impact of the knowledge described within the publication. While the full range of citation analysis issues (and the extensive literature involved in these issues) is much too large to be summarized here, the analysis aims to implement a state-of-the-art empirical approach incorporating key insights from this literature (See CHI Research, 2002; Jaffe and Trajtenberg, 2002; Adams and Griliches, 1996; and Moed et al,1986).

## CHAPTER 4

### BOX 4B

#### CHOOSING THE SAMPLE

The data in our analysis consist of two subsamples. For our “base” dataset, we have assembled a collection of research articles linked to BRC deposits, along with “matched” samples of control articles. To build the sample of BRC-linked articles, we take advantage of the fact that ATCC prepares reference information for materials deposited within its collections. For each ATCC deposit, the ATCC catalog (maintained online at [www.ATCC.org](http://www.ATCC.org), and historically published in catalog-form) identifies the references associated with ATCC deposits, the name of the original depositor, the date of the deposit, and key scientific articles associated with the deposit. For each deposit, we consider the first article listed within the ATCC deposit reference section as the “focal” article for that deposit.<sup>4</sup> These articles are associated with 183 deposits randomly selected from the materials deposited among three of ATCC’s primary collections (Bacteria, Cell Biology, and Molecular Biology) between 1984 and 1999.

For our more nuanced *differences-in-differences* analysis, we employ a second subsample, the “Special Collections” data. This sample is composed of 127 articles linked to materials in three special collections that have been transferred in bulk to the ATCC from private culture collections.<sup>5</sup> While the first subsample allows us to evaluate a broad cross-section of

---

<sup>4</sup> Multiple members of the scientific and information technology staff at ATCC with whom we conducted interviews suggest that the first reference article is typically the one most closely associated with initial use of the biological material.

<sup>5</sup> Numerous scientists, research institutions, and corporations maintain private collections. With the exception of those collections operated by firms, many of these allow open access to their collections; on balance, however, they are less engaged in characterization and knowledge of the contents of their collections is less well-diffused.

BRC-linked articles, the special collections subsample allows us to disentangle the intrinsic impact of the research quality from the impact that results from deposit in a BRC.

Each of the BRC-linked articles is matched with two types of control articles. Each type of control article is designed to be as similar to the ATCC-associated article on as many observable dimensions as possible. In the first set of controls, the “nearest neighbor” controls, we use the PUBMED database to match each BRC-linked article with the article that immediately precedes it in the same issue of the same journal. For example, if an ATCC-associated reference were the third article in the June 14, 1986 issue of *Journal of Cell Biology*, the *nearest neighbor* control article would be the second article within that same issue.<sup>6</sup> This ensures that both the BRC-linked and control article have undergone the same scientific review process and been published at the same moment in time. Because some journals cover a wide range of scientific disciplines – e.g., *Science* and *Nature* – we also collect citation data for another type of control. Using a National Library of Medicine algorithm, we identify for each BRC-linked article, the “most related” article that appears in the same volume of the same journal. For example, if an ATCC-associated reference were published in the June 1986 issue of *Cell*, the *most-related article control* would be the article in the 1986 edition of *Cell* whose research topics and themes were judged by the National Library of Medicine (based on keywords, title, and abstract) to be the one most similar to the ATCC-associated article.

We then use the Science Citation Index (as well as other bibliometric information) to assess the relative impact of these articles on subsequent scientific research. In particular, as described in the main text, we compare the citation patterns of articles associated with ATCC

---

<sup>6</sup> In the event that the ATCC-associated article is the lead article in its particular issue, we use the second article in that journal as the control.

material deposits to the citation patterns of control articles. In so doing, we are able to identify the impact of BRC material deposits.

## CHAPTER 4

### BOX 4C

#### CITATION ANALYSIS DATA AND MEASURES

The empirical analysis is based on a dataset composed of 127 BRC-linked articles and accompanying control articles (See Box 4B, “Choosing the Sample” for a description). For each of these articles, we measure citations for each year beginning in the year of publication through December, 2001. Table 4C-1 provides the summary statistics for the key measures. Citation activity is measured as Forward Citations, the annual number of citations received by an article, and Cumulative Citations, the summation of citations received through the end of 2001. Overall, the level of citation by articles in the sample is quite high (compared to a completely random sample of life sciences publications). This high citation rate is not surprising, since articles associated with BRC deposits tend to be published in top-tier scientific journals such as *Science*, *Nature*, or *Cell*. Similar to other citation datasets, the distribution is quite skewed (with many observations of Forward Citations equal to zero and a small number of outliers with over a hundred citations in a single year).

The analysis includes additional measures for the timing of publication (PUBLICATION YEAR), the timing of deposit (DEPOSIT YEAR), and the length of time lapsed since initial deposit (AGE).<sup>7</sup> While the date of publication varies from 1970 through 2001, all accession dates in this analysis occur from 1981 onwards (this is the first year we drew articles for the random sample and the first accession date for articles associated with the special collections). We have also investigated the impact of several additional article characteristics, including the

---

<sup>7</sup> In some cases, the DEPOSIT YEAR is measured with error (of up to a few months). As materials moved wholesale into ATCC must undergo authentication and cataloging before they are available to public use, there is some delay between the announcement of a transfer and ATCC’s ability to ship materials for scientific use. In some occasions, materials may be available for a few months before their accession is officially declared in a catalog or other ATCC publication.

number of pages for each article (# PAGES), the number of authors (# AUTHORS), the number of backward citations (BACKWARD CITATIONS). As well, we collected characteristics of the lead author of each article including their institutional affiliation, whether this affiliation was Companies, University or Government, and whether the location was Foreign. Finally, for articles linked to BRC deposits, we collect the PRICE charged by the ATCC for access, which averages approximately \$230 over the sample.

## CHAPTER 4

### BOX 4D

#### THE SPECIAL COLLECTIONS<sup>8</sup>

The second stage of the empirical analysis takes advantage of the fact that “orphan” collections are occasionally transferred in bulk from non-BRC private collections. Such transfers usually occur as the result of funding crises or retirement of key personnel from the institution in which the collection is stored. Of the special collections maintained by the ATCC, three were particularly interesting, since their inclusion in the ATCC collections was not directly related to an increased perception of their scientific importance. The first is a set of articles associated with the Gazdar Collection. This collection was transferred into the ATCC when Dr. Adi Gazdar left his position as Head of Tumor Cell Biology Section at the National Cancer Institutes, along with his collaborator, Dr. John Minna, to become Professor of Pathology at the Hamon Center for Therapeutic Oncology at UT Southwestern. The Gazdar collection was incorporated into ATCC over a number of years; the materials examined in this paper were accessioned beginning in 1994. The second set of materials is drawn from the Tumor Immunology Bank (TIB), which was transferred from the Salk Institute in 1981 due to funding considerations and was accessioned beginning in 1982. The final set of articles in the dataset is associated with materials in the Human Tumor Bank (HTB). The HTB had been maintained by researchers at Sloan-Kettering until funding considerations led to its being transferred into ATCC beginning in 1981.

---

<sup>8</sup> Historical details on ATCC’s collections are drawn from discussions with Dr. Robert Hay, director of the Department of Cell Biology at ATCC.

## CHAPTER 4

### BOX 4E

#### THE RESEARCH METHODOLOGY

Two key challenges are associated with isolating out the distinct impact that BRC deposit has on Forward Citations. First, it is difficult to disentangle the marginal impact of BRC deposit from the possibility that materials ultimately deposited in BRCs are associated with articles of higher intrinsic scientific importance. We address this challenge by implementing a *differences-in-differences* empirical approach (Angrist, Imbens, and Rubin, 1996). where the estimate of the impact of BRCs is inferred from the *change* in citation patterns that result after BRC deposit (relative to the trend that would be experienced in the absence of deposit). By simultaneously comparing citation patterns *across article pairs* (i.e., comparing articles eventually deposited in BRCs with those that are not) and *across deposit-status within article* (i.e., whether a particular article has yet been deposited), we can separately identify the degree to which articles have higher intrinsic importance (i.e., the selection effect) and the degree to which ATCC deposit has a marginal impact on citation rates. Second, because citation data are realized in the form of annual “count” data and are highly skewed to the right (i.e., the median is substantially smaller than the mean), the use of a traditional regression techniques, such as Ordinary Least Squares (OLS), are inappropriate. Instead, building on a decade of research in this area, we employ a negative binomial regression model, which accounts for both count data and the skewness of citation patterns (see Appendix 4A; Cameron and Triverdi (1998); and Furman and Stern (2003), for further discussion).

## E. References

- Adams, J. and Z. Griliches (1996). "Measuring Science: An Exploration," Proceedings of the National Academy of Sciences, 93: 12664-70.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). *Identification of Causal Effects using Instrumental Variables*. Journal of the American Statistical Association 91: 444--72.
- Cameron, A.C. and P.K. Triverdi (1998). *Regression Analysis of Count Data*. Cambridge University Press.
- CHI Research (2002). [www.chirsearch.com](http://www.chirsearch.com)
- OECD (2001). *Biological Resource Centres: Underpinning the Future of Life Sciences and Biotechnology*. Paris, France: OECD Books.
- Furman, J. and S. Stern (2003). "Climbing Atop the Shoulders of Giants: Institutions The Impact of Institutions on Cumulative Research," mimeo, Northwestern University.
- Griliches, Z., ed. (1984). *R&D, Patents and Productivity*. Chicago (IL): Chicago University Press.
- Griliches, Z. (1990). "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 92, 630-653.
- Griliches, Z. (1998) *R&D and Productivity: The Econometric Evidence*. Chicago (IL): University of Chicago Press.
- Jaffe, A., M. Trajtenberg, and R. Henderson (1993), Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *The Quarterly Journal of Economics*, 434, 577-598.

Moed, H.F. et al (1985). "The Use of Bibliometric Data for the Measurement of University Research Performance," *Research Policy* 14: 76-X.

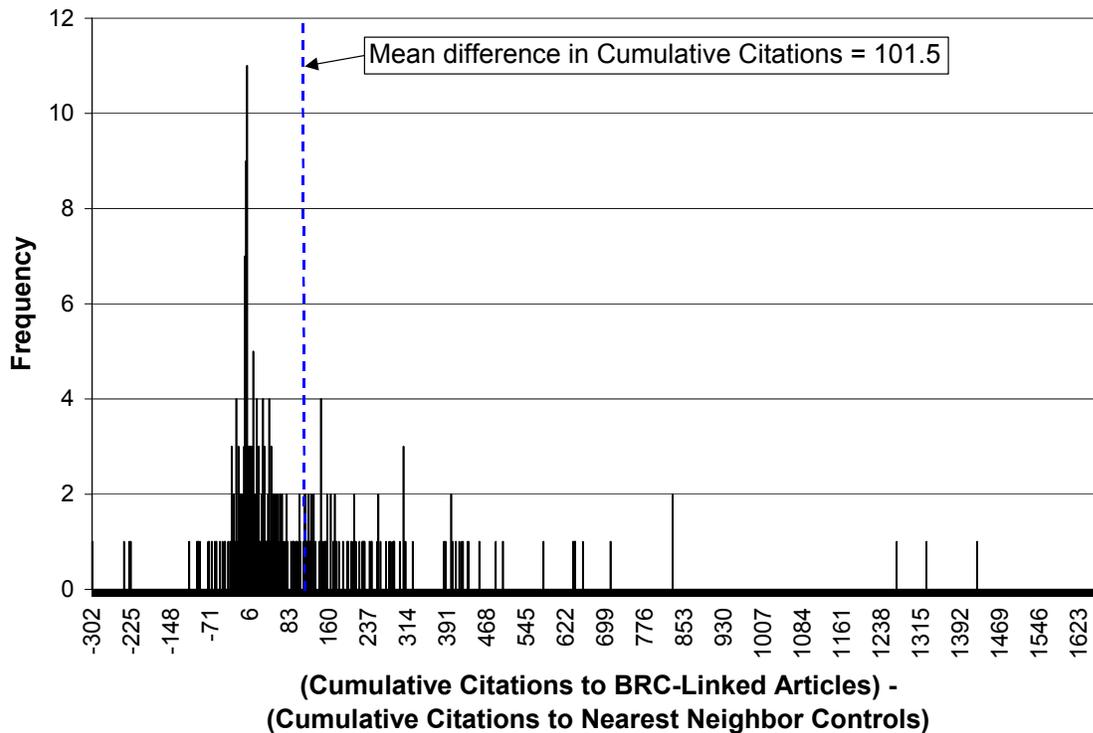
National Science Foundation (2002). *Science and Engineering Indicators*.

Wang, P., I. Cockburn, and M. Puterman "Analysis of Patent Data - A Mixed Poisson Regression Approach," *Journal of Business and Economic Statistics*, 1998, 16(1), pp. 27-41.

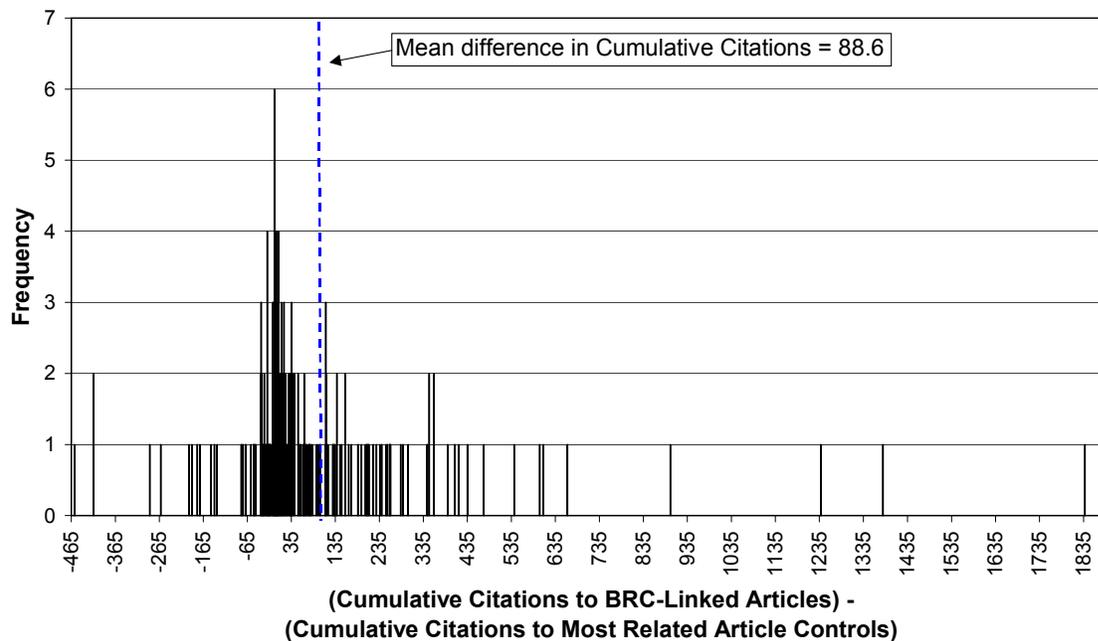
**FIGURE 4A**  
**DIFFERENCE IN CITATION RATES,**  
**BRC-LINKED ARTICLES VS. CONTROLS**

	Average for ATCC Sample	Average for "Nearest Neighbor Article" Controls	%Difference: ATCC / (Nearest Neighbor Article Controls)	Average for "Most-Related Article" Controls	%Difference: ATCC / (Most-Related Article Controls)
Total Backward Citations	32.9	26.4	125%	29.7	111%
Total Forward Citations (through 2001)	151.0	49.5	305%	79.2	191%
Average Forward Citations Per Year	9.2	3.0	305%	4.7	196%

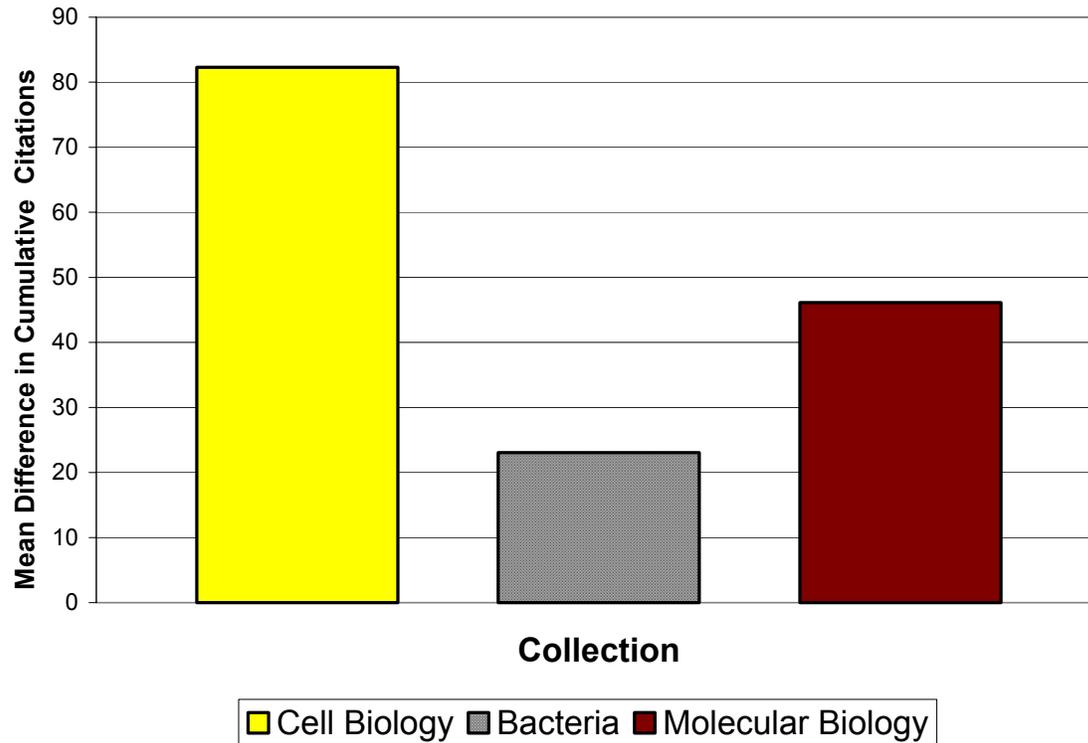
**FIGURE 4B-1**  
**DIFFERENCE IN CUMULATIVE CITATIONS,**  
**BRC-LINKED ARTICLES VS. NEAREST NEIGHBOR CONTROLS**



**FIGURE 4B-2**  
**DIFFERENCE IN CUMULATIVE CITATIONS,**  
**BRC-LINKED ARTICLES VS. MOST-RELATED ARTICLE CONTROLS**

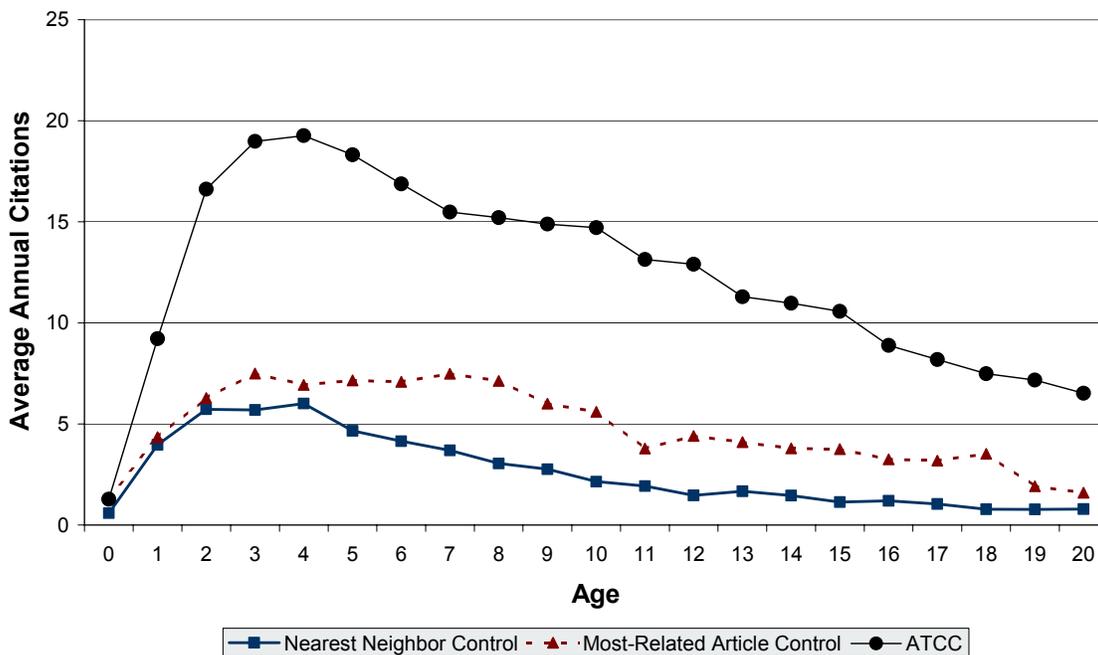


**FIGURE 4C**  
**MEAN DIFFERENCE IN CUMULATIVE CITATIONS BETWEEN**  
**BRC-LINKED ARTICLES AND NEAREST NEIGHBOR CONTROLS,**  
**BY COLLECTION\***

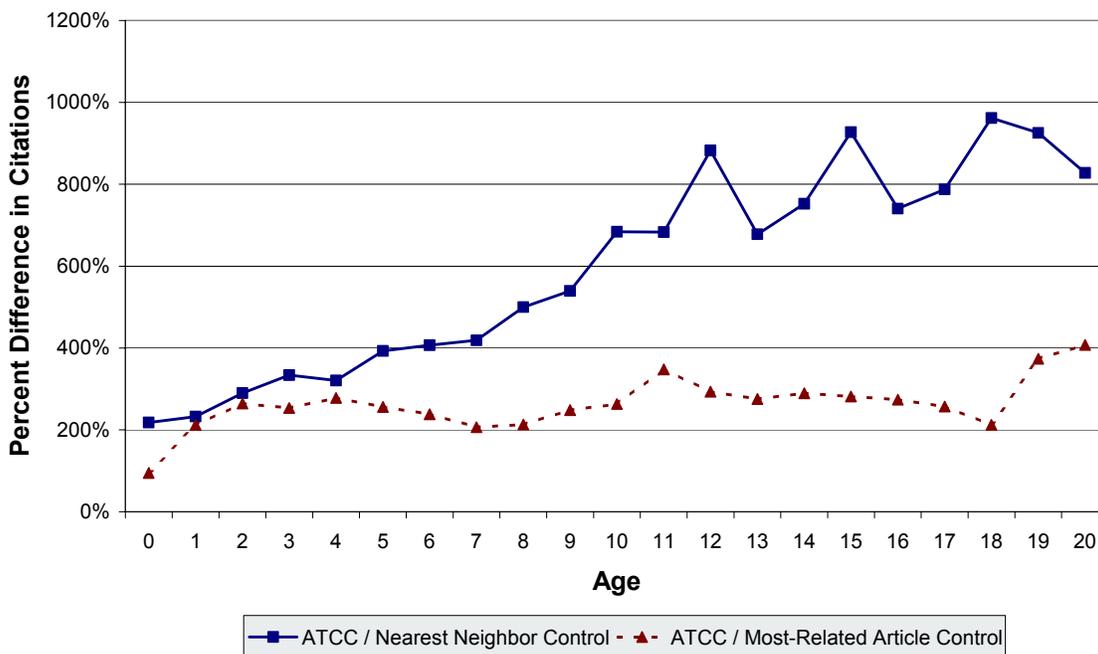


\* Note: This figure is based on a subset of the original dataset, for which the Collection (Cell Biology, Bacteria, or Molecular Biology) is identified.

**FIGURE 4D-1  
AVERAGE ANNUAL CITATIONS BY AGE,  
BRC-LINKED ARTICLES VS. CONTROLS**



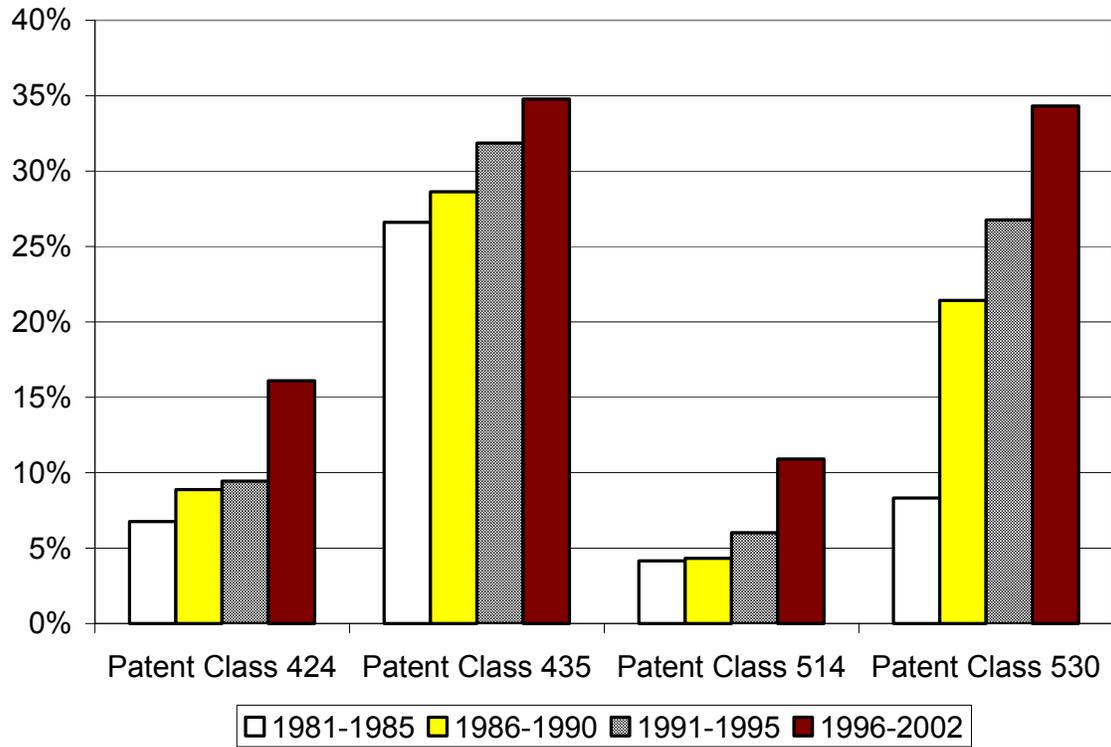
**FIGURE 4D-2  
PERCENT DIFFERENCE IN ANNUAL AVERAGE CITATIONS TO  
BRC-LINKED ARTICLES VS. CONTROLS, BY AGE**



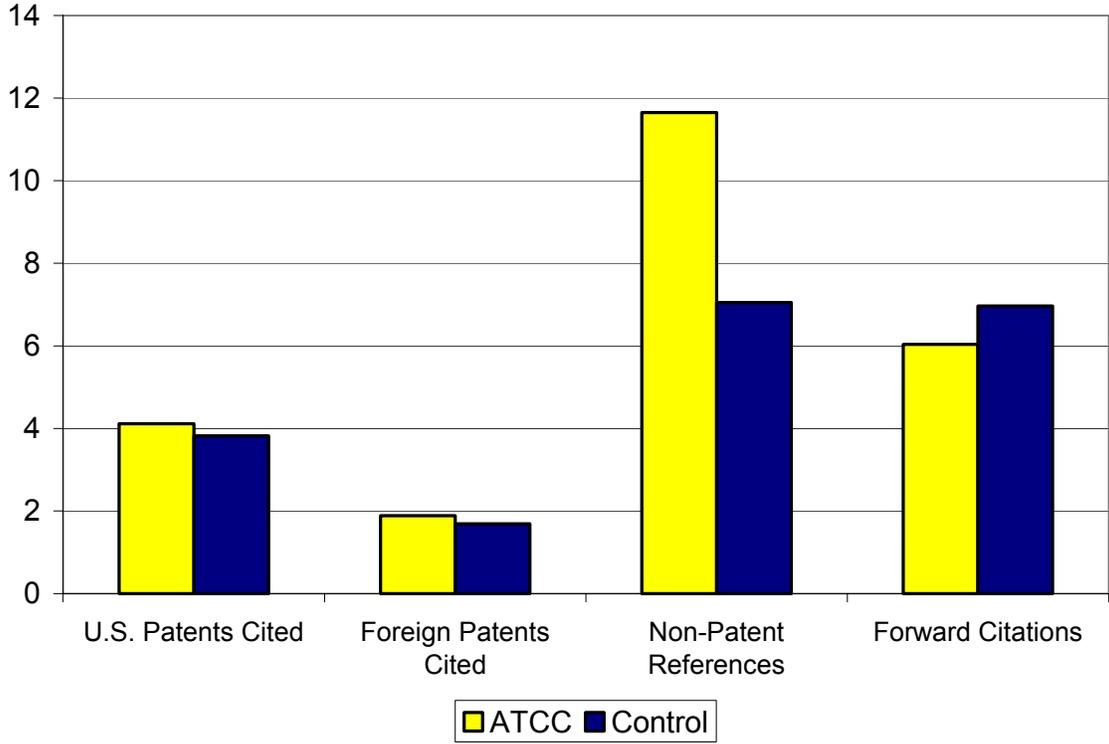
**FIGURE 4E**  
**DIFFERENCES IN ARTICLE CHARACTERISTICS,**  
**BRC-LINKED ARTICLES VS. CONTROLS**

	<b>BRC-LINKED ARTICLES</b>	<b>NEAREST NEIGHBOR ARTICLE CONTROL</b>	<b>MOST RELATED ARTICLE CONTROL</b>
NUMBER OF PAGES	8.1	6.8	4.3
NUMBER OF AUTHORS	4.2	3.7	7.7
% UNIVERSITY AUTHORS	57.5%	60.3%	60.2%
% GOVT. AUTHORS	14.7%	13.1%	11.8%
% FOREIGN AUTHORS	25.3%	40.7%	39.1%

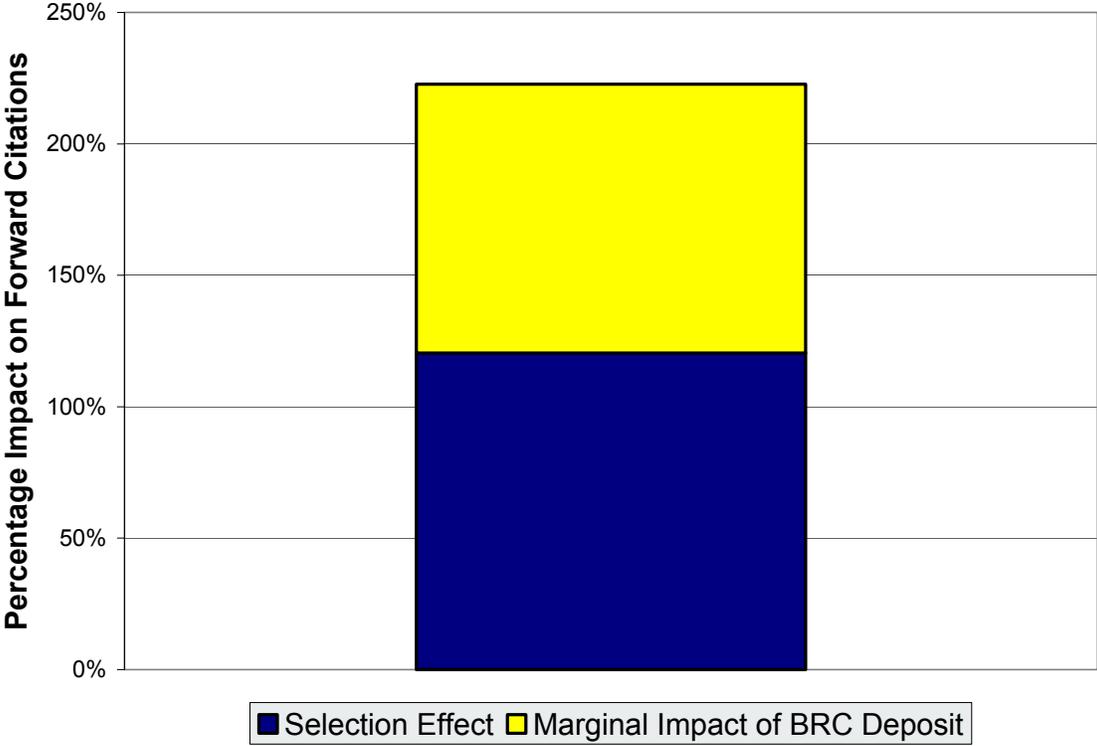
**FIGURE 4F**  
**FRACTION OF PATENTS REFERENCING ATCC IN**  
**KEY PATENT CATEGORIES, BY TIME PERIOD**



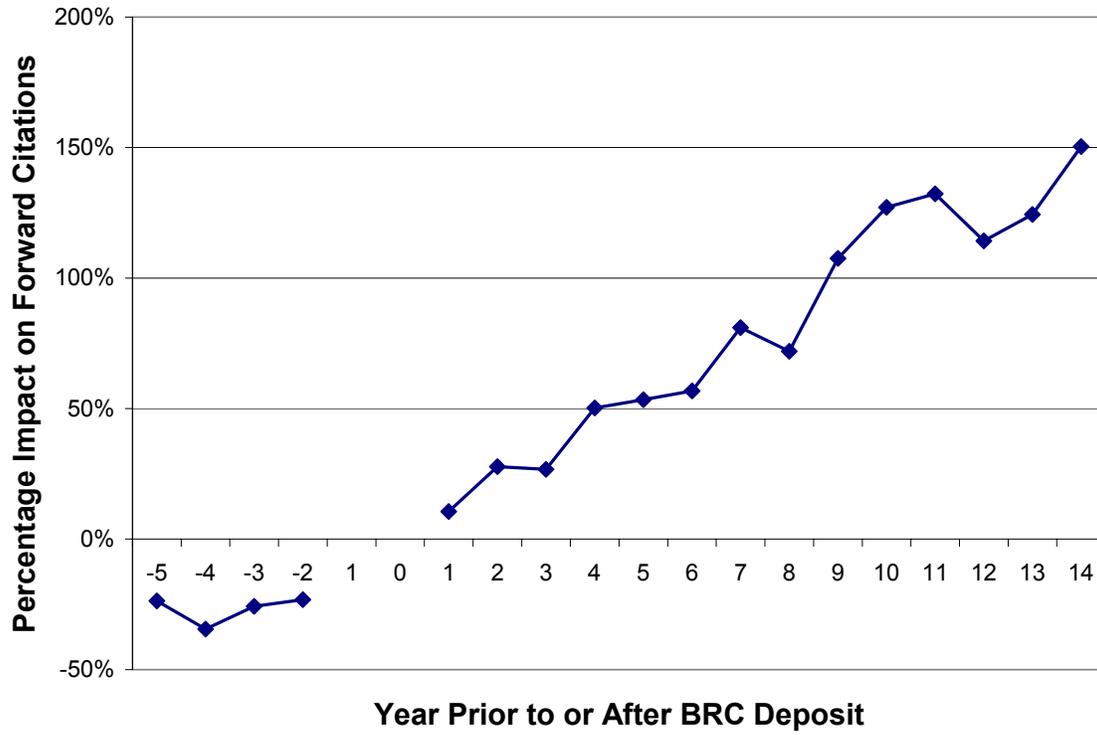
**FIGURE 4G**  
**DIFFERENCES IN CHARACTERISTICS,**  
**ATCC-REFERENCING PATENTS VS. CONTROLS**



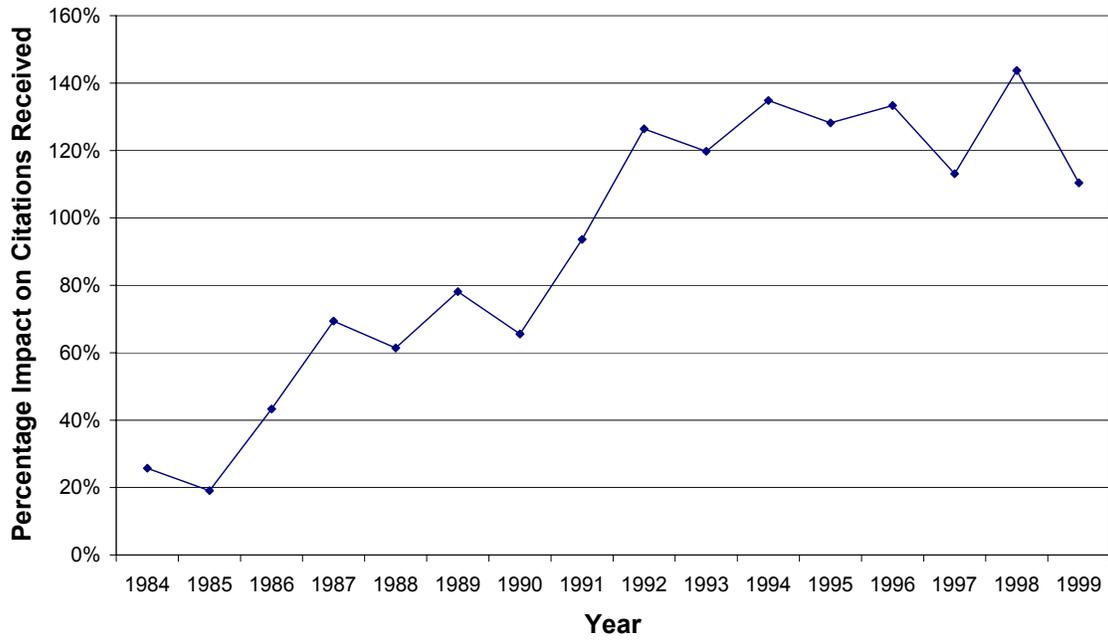
**FIGURE 4H**  
**SELECTION EFFECT VS. MARGINAL IMPACT OF BRC DEPOSIT**



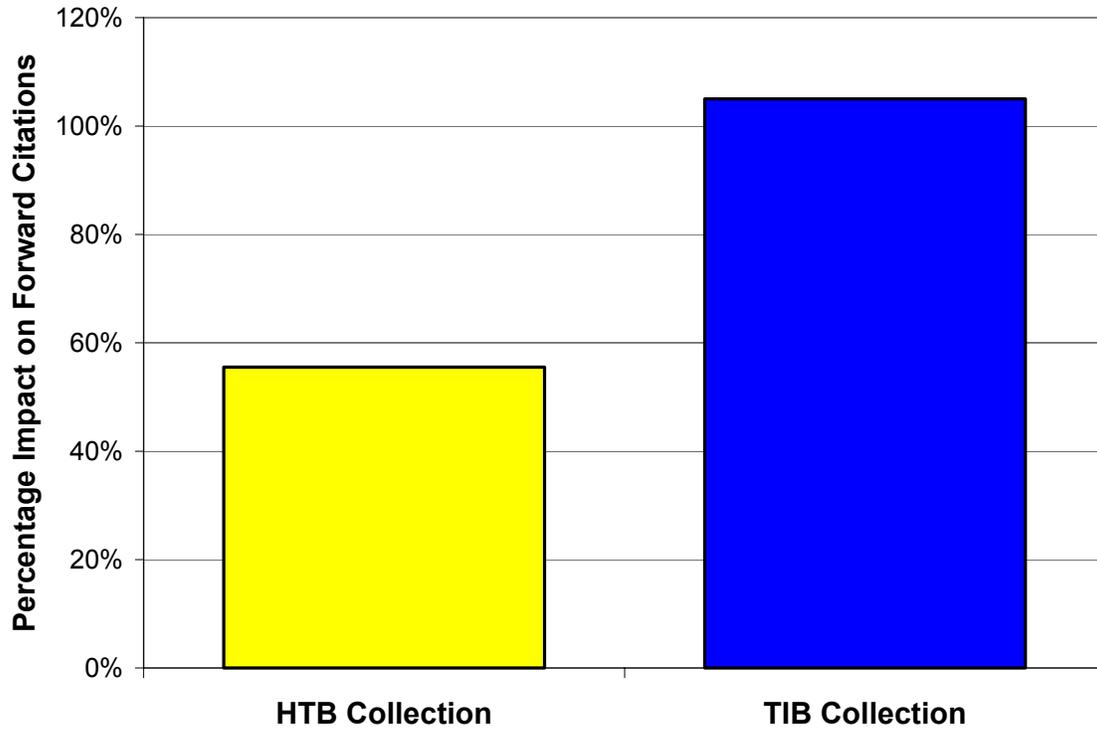
**FIGURE 4I**  
**IMPACT OF BRC DEPOSIT ON FORWARD CITATIONS**  
**IN YEARS PRIOR TO AND AFTER BRC DEPOSIT**



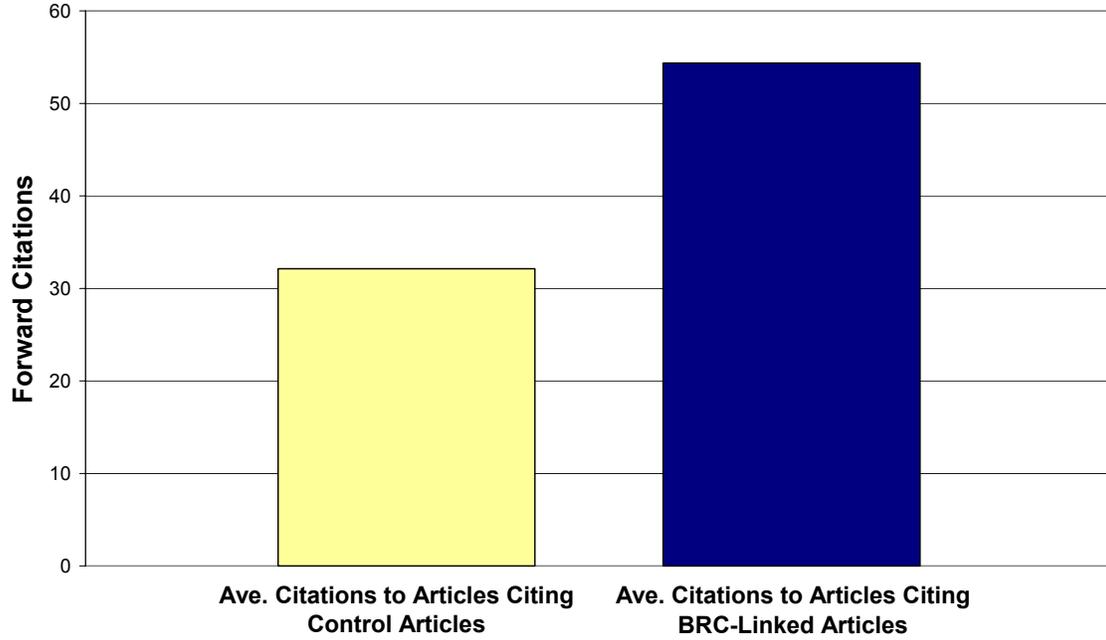
**FIGURE 4J**  
**IMPACT OF BRC DEPOSIT ON FORWARD CITATIONS,**  
**MARGINAL EFFECTS BY YEAR**



**FIGURE 4K**  
**IMPACT OF ATCC DEPOSIT ON FORWARD CITATIONS,**  
**BY “SPECIAL COLLECTION”**



**FIGURE 4L**  
**AVERAGE FORWARD CITATION RATES**  
**FOR ARTICLES THAT CITE BRC-LINKED AND**  
**NEAREST NEIGHBOR CONTROL ARTICLES**



**FIGURE 4M**  
**BRC DEPOSIT COST-EFFECTIVENESS ANALYSIS**

<b>Estimated Baseline Citation Cost*</b>	<b>BRC Accession Cost^</b>	<b>Article category</b>	<b>BRC Citation Boost</b>	<b>BRC Citation Cost</b>	<b>BRC Cost-Effectiveness Index</b>
\$2,400	\$10,000	<b>Articles published by researchers in “Top Ten” Biology Departments</b>	20.17	\$459.7	4.84
		<b>Articles published by researchers in Random Biology Departments</b>	11.73	\$852.8	2.81
		<b>Average articles associated with BRC deposits</b>	11.14	\$898.7	2.67

\* Estimated baseline citation cost drawn from Adams & Griliches (1996).

^ Estimated BRC accession cost drawn from OECD (2001).

**TABLE 4C-1  
SUMMARY STATISTICS**

<b>VARIABLE</b>	<b>N</b>	<b>MEAN</b>	<b>STANDARD DEVIATION</b>	<b>MIN</b>	<b>MAX</b>
<b>ARTICLE-YEAR MEASURES</b>					
FORWARD CITATIONS	13947	5.77	12.65	0	186
BRC-LINKED	13947	0.38	0.48	0	1
YEAR	13947	1991.91	6.56	1970	2001
AGE	13947	9.10	6.56	0	31
<b>ARTICLE DEPOSIT &amp; AUTHOR CHARACTERISTICS</b>					
PUBLICATION YEAR	844	1985.47	6.64	1970	1999
# PAGES	842	6.69	5.59	0	70
# AUTHORS	838	4.86	3.69	0	57
UNIVERSITY	772	0.59	0.49	0	1
GOVERNMENT	772	0.13	0.34	0	1
FOREIGN	752	0.34	0.48	0	1
<b>ARTICLE CHARACTERISTICS*</b>					
DEPOSIT YEAR*	127	1983.56	3.35	1981	1994
ATCC PRICE*	127	229.80	44.01	167	270

\* DEPOSIT YEAR and PRICE data exist only for those associated with deposits to the ATCC Special Collections. All other data are from the Base Sample.

## CHAPTER 4

### APPENDIX 4A

This brief appendix discusses the citation regression findings in greater detail. As described in Box 4C and Furman and Stern (2003), the data consist of 127 sets of BRC-linked “Special Collections” articles along with two sets of matched control articles. The data include citation data as well as article, deposit, and author characteristics. Because citation data are realized in the form of annual “count” data and are highly skewed to the right (i.e., the median is substantially smaller than the mean), the use of a traditional linear regression model (such as OLS) is inappropriate. Estimates will be downward-biased as the analysis will “overweight” the high prevalence of observations for which the number of annual citations is equal to zero. While a Poisson estimation is the most traditional approach for dealing with count data, research over the past two decades suggests that the strong restrictions of the Poisson model (specifically, that the mean and variance of the underlying count data distribution are equal) can yield misleading results (Wang, Cockburn, and Puterman, 1998). Instead, an appropriate (and commonly used) specification is the negative binomial regression model (a Poisson model in which the variance differs from the mean). For a complete discussion of the advantages of the negative binomial in this context, see Cameron and Triverdi (1998) and Furman and Stern (2003).

Table A-1 presents the core findings. For each column, the dependent variable is annual FORWARD CITATIONS, and each specification includes year fixed effects, vintage fixed effects, and either article-pair or article fixed effects (Furman and Stern (2003) includes a complete set of specifications, some of which exclude the control variables). In addition to the variables defined earlier in the text, we define a new variable, BRC-LINKED, POST-DEPOSIT

which is only equal to one for those article-years after the material associated with the article have been deposited at the ATCC. (4A-1) and (4A-2) regressions include both BRC-LINKED and BRC-LINKED, POST-DEPOSIT, allowing us to separate out the selection effect (i.e., BRC-LINKED captures the degree to which articles have higher citation rates independent of whether they have been deposited) from the marginal impact of BRC deposit (captured by the coefficient on BRC-LINKED, POST-DEPOSIT).

(4A-1) presents an OLS specification of FORWARD CITATIONS on BRC-LINKED and BRC-LINKED, POST-DEPOSIT with control variables included. The results are significant: According to the OLS specification, articles associated with BRC-deposited materials receive 57.7% more annual citations; after those materials are accessioned into ATCC, associated articles receive an 46.5% boost in citations. This suggests that important articles are “selected” into BRCs and that BRCs have an important impact in enhancing the diffusion of knowledge associated with these articles.

As discussed earlier, however, OLS is inappropriate for count data. We therefore turn to negative binomial specifications. For these regressions, coefficients are reported as incidence-rate ratios (i.e., a coefficient equal to one implies no effect while a 1.50 coefficient implies a 50% boost). The results in (4A-2) confirm with a negative binomial specification what (4A-1) suggests with OLS. Both the selection effect and the marginal impact of BRC-deposit are statistically and economically significant – the findings show that BRC-linked articles have double propensity to be cited than control articles.

The final two columns report the core findings associated with the differences-in-differences analysis. In these models, we omit the BRC-LINKED dummy and employ article fixed effects (rather than article pair fixed effects) to control for article-specific heterogeneity.

The coefficient BRC-LINKED, POST-DEPOSIT again reflects the post-deposit citation boost experienced by BRC-linked articles. In these models, as well, the accession of materials into ATCC is demonstrated to have an important impact on associated articles. Controlling for all time-invariant article-specific characteristics, ATCC deposit is associated with an 82 percent boost in citations. Expanding on the results in the earlier analysis, (4A-4) demonstrates that the impact of deposit increases over time, as well. Overall, the econometric evidence provides support for a model in which biological resource centers play a key role in the cumulative capacity of the life sciences.

**TABLE 4A-1**  
**TIME-SERIES RESULTS ON “EXPERIMENTAL DATA” ONLY\***

	OLS Dep Var = ln(FORWARD CITATIONS-1)	NEGATIVE BINOMIAL (Coefficients reported as incidence-rate ratios) Dep Var = FORWARD CITATIONS		
	(4A-1) Overall BRC Effect*	(4A-2) Baseline Count Model*	(4A-3) Core Diff-in-Diffs Model	(4A-4) (A-3) Interaction between ATCC Article * Time Since Deposit
<b>ARTICLE CHARACTERISTICS</b>				
BRC-LINKED	<b>0.577</b> (0.050)	<b>2.204</b> (0.303)		
BRC-LINKED, POST- DEPOSIT	<b>0.465</b> (0.054)	<b>2.023</b> (0.277)	<b>1.824</b> (0.199)	<b>1.677</b> (0.166)
BRC-LINKED * TIME SINCE DEPOSIT				<b>1.025</b> (0.012)
<b>CONTROL VARIABLES</b>				
<i>Parametric Restrictions</i>				
Article Pair FEs = 0	# Restrict 126 F-stat 4.7*10 <sup>5</sup> p-value 0.00	# Restrict 65 $\chi^2$ 3.6*10 <sup>11</sup> p-value 0.00		
Article FEs = 0			# Restrict 63 F-stat 1.4*10 <sup>7</sup> p-value 0.00	# Restrict 64 F-stat 6.8*10 <sup>6</sup> p-value 0.00
Age FEs = 0	# Restrict 30 F-stat 20.98 p-value 0.00	# Restrict 30 $\chi^2$ 265.66 p-value 0.00	# Restrict 30 F-stat 417.85 p-value 0.00	# Restrict 30 $\chi^2$ 418.83 p-value 0.00
Year FEs = 0 <sup>~</sup>	# Restrict 23 F-stat 2.24 p-value 0.00	# Restrict 23 $\chi^2$ 58.94 p-value 0.00	# Restrict 23 F-stat 67.95 p-value 0.00	# Restrict 23 $\chi^2$ 67.37 p-value 0.00
<b>REGRESSION STATISTICS</b>				
R-squared	0.52			
Log-likelihood		-16246.01	-14396.54	-14385.79
# of Observations	7198	7198	7198	7198

\* Robust standard errors are in parentheses.

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

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