Materials Transfer Agreements (MTAs), Licenses, and the Flow of Scientific Knowledge[1]

Neil C. Thompson
Haas School of Business
U.C. Berkeley

David C. Mowery
Haas School of Business
U.C. Berkeley
and NBER

Arvids A. Ziedonis
Lundquist College of Business
University of Oregon

PRELIMINARY: DO NOT CITE OR QUOTE WITHOUT AUTHORS' PERMISSION

ABSTRACT

As university involvement in "technology transfer" and entrepreneurship has grown, concerns have been raised about the impact of academic patenting on academic science. Murray and Stern (2007) suggest that scientific publications in the field of biotechnology experience a decline in citations following the issue of a patent on that research. They interpret this decline as evidence that patenting may have a "chilling effect" on scientific communications, as measured by citations to published academic papers. Other authors, however, notably Walsh et al. (2005) suggest that academic scientists pay little attention to patents in developing their research agenda. Walsh et al. (2007) argue that in some fields, Material Transfer Agreements (MTAs) may constrain the development of scientific research.

This paper extends the work of Murray and Stern to cover a broader sample of published scientific papers in the biomedical and other disciplines, and examines the effects of MTAs and licenses on citations to these papers. We utilize data on invention disclosures, patents, and MTAs from the University of California system during 1997-2007, and employ a new technique for matching publications to patents, MTAs, and licenses.

We find that *in general* MTAs and licenses have little impact on citations to related published articles. We do find some significant negative effects on citations for publications associated with MTAs with the private sector, and licenses on research tools. We also find evidence of a larger positive effect of licenses for publication citations since 2002.

---

[1] We are grateful to Wendy Streitz, Ron Franck, and other staff employees of the UC Office of the President for invaluable assistance in gaining access to the intellectual property data employed in this paper, and to Thompson Reuters for their cooperation in obtaining the data on scientific publications. We also thank Professors Fiona Murray and Scott Stern of the Sloan School at MIT for sharing their data with us. Earlier versions of this paper benefited from the comments of participants in seminars at the Haas School of Business, U.C. Berkeley, especially those of Professors Brian Wright and Bronwyn Hall. The authors also gratefully acknowledge financial support from the Institute for Business Innovation at the Haas School of Business, the Industrial Partnerships Office at Lawrence Livermore National Laboratory, and the Program in Law, Economics, and Technology at the University of Michigan Law School.

I. Introduction: The effects of patenting on scientific communication

In his influential book on science and technology policy, Stokes (1997) argued that the conventional distinction between basic and applied research overlooked an important intermediate class of research, research that sought fundamental scientific understanding that was also motivated by considerations of utility. Research in "Pasteur's Quadrant," according to Stokes, produces fundamental knowledge and in many cases, artifacts with significant commercial applications. Both universities and industrial research facilities (e.g., Bell Labs, which supported an extensive program of "mission-oriented basic research") have long undertaken this type of work. Moreover, the results of such research long have been disclosed publicly through publications and have generated patents by both industrial firms and U.S. universities.

Although Pasteur's Quadrant is not a new realm for scientific research in U.S. universities, an array of factors, including the Bayh-Dole Act of 1980 and related changes in U.S. intellectual property rights law and policy, as well as expanded federal support for academic biomedical research, have been associated with increased patenting of the results of academic research. Previous work on this topic has examined the effects of patenting on biomedical researchers' willingness to share information on their work (Blumenthal et al., 1997; Campbell et al., 2002). More recent research has analyzed the effects of patenting biomedical discoveries that are also disclosed in scientific papers. Some of this work finds that the issuance of a patent results in modest but significant declines in citations to the research papers related to the patent (Murray and Stern, 2007; Sampat, 2005). Other research, however, argues that biomedical researchers rarely if ever search to determine whether a prospective research project or experiment will infringe on patents (Walsh, et al., 2005; Lei et al., 2009). In a follow-on study, Murray and Stern (2008) argue that the life sciences community adapted to patenting of academic research by the end of the 1990s to such an extent that by the early 2000s scientists "no longer considered patents salient" (p. 36).

This paper extends the work of these scholars in several ways. Using data from the University of California system, we examine the effects of agreements governing the exchange of research materials (Materials Transfer Agreements, or MTAs) and of licenses on the citations received by scientific papers that were also patented. In addition, we expand the universe of scientific disciplines covered by this work to include non-biomedical fields such as the physical sciences and engineering.

Immediately below, we discuss the potential effects of MTAs and licenses on the flow of scientific knowledge among researchers. We then discuss our data and estimation approach, followed by a discussion of preliminary results and the robustness of those results. We conclude with a discussion of the implications of our findings and our future research agenda.

II. Materials Transfer Agreements (MTAs) and Scientific Communication

MTAs are agreements among researchers governing the transfer and exchange of materials, usually biological, used in research. Their detail and complexity vary, but many include provisions for royalties on patents resulting from the use of the materials or limit the ability of the recipient to patent or license the results of research that uses the materials. MTAs are used widely by both industry and academic researchers, and cover exchanges of materials within industry, within academia, and between industry and academia.

The exchange by researchers of biological materials for use in fundamental research has a long and occasionally controversial history in the biomedical sciences.[2] Historically, materials exchanges were governed by little more than a letter from the source accompanying the materials, requesting acknowledgement and in some cases asking that the materials not be passed on to third parties (See McCain, 1990). The more elaborate MTAs used in contemporary materials exchanges appear to be a byproduct of the post-1980 surge in academic patenting.[3] Many MTAs used for exchanges of materials among academic researchers, especially those governing materials exchanges between industrial and academic researchers, now contain clauses requiring that the recipient of the materials surrender all claims to intellectual property based on discoveries using the materials (Marshall, 1997). In other cases, the source of the materials being requested has required a royalty on any commercial product resulting from research employing the material, a so-called "reach-through licensing agreement" (RTLA). Even if MTAs' provisions do not limit researchers' freedom to pursue a particular agenda, access to important tools or materials may be impeded or delayed by the complexity of negotiating them or of obtaining approval by academic or industrial administrators or managers.

The overall increase in academic patenting of biomedical discoveries, as well as the higher perceived value of biological and genomic materials used in biomedical research, has expanded the number and diversity of the institutions seeking to obtain or being asked to provide these materials to other researchers. The greater diversity of participants makes the negotiation of satisfactory terms among the parties to a given MTA more difficult, according to Eisenberg (2001),[4] and thus can delay researcher

---

[2] One celebrated controversy concerned the failure of the Gallo research team at the NIH to acknowledge that the pathbreaking isolation of the AIDS virus relied on a cell line established by another research team, at the Veterans Administration Clinical Oncology Branch (See Rubinstein, 1990 for further details). The Milstein-Kohler hybridoma technique for producing cell lines also was patented not by the discoverers but by another research team that requested and received a sample of the Milstein laboratory's plasmacytoma cells (See Wade, 1980).

[3] Respondents to the survey of University of California agricultural biotechnology researchers by Lei et al. (2009) report "moderately more" use of MTAs than in 1999.

[4] The implications of a more diverse array of parties to these exchanges are complex, as the NIH Working Group on Research Tools (chaired by Professor Eisenberg) pointed out in its 1998 report: "The very term 'research tool' connotes a user perspective

access to materials. Complex MTAs appear to be more common in materials exchanges that span the academia-industry divide.[5] Still another problem associated with the growing use of MTAs and their escalating complexity is the demands on licensing office staff for review and approval. The director of the University of Pennsylvania technology licensing office noted in 1997 that that number of MTAs reviewed by his office had more than doubled during the previous 12 months from 197 to 425, even as the provisions of many of them had become more complex (Marshall, 1997). The NIH Working Group on Research Tools reported that the University of Washington's technology licensing office was handling an average annual volume of "incoming" MTAs (requested by other institutions' researchers) in the mid-1990s of roughly 1,000.

A different perspective on the role and effects of MTAs is provided by Stern (2004) in his discussion of biological resource centers, as well as by Walsh et al. (2003). Biological resource centers (BRCs) are nonprofit materials depositories that play a key role in maintaining the reliability and provenance of cell lines used by industrial and academic researchers—as Stern notes, contamination of widely used cell lines has caused major research fiascoes in the past several decades. Stern argues that the use of MTAs by BRCs has aided the exchange of materials,[6] and recommends that MTAs be a standard complement to patents covering biological discoveries: "Putting MTAs in place at the time of patent approval lowers the cost of mutually beneficial transactions between the developers of materials and follow-on researchers and widens the availability of patented biomaterials." (2004, pp. 96-97). Similarly, Walsh et al. (2003) argue that the formalization of materials exchanges through MTAs may simplify these transactions and facilitate researcher access.[7]

A more recent paper by Walsh et al. (2007) surveys biomedical researchers on the extent to which patents and MTAs constrain their research activities. Consistent with their earlier paper, Walsh et al. (2005), researchers reported that patents on relevant intellectual property did not significantly limit their

---

rather than a provider perspective. What a user sees as a research tool, a provider may see as a valuable end product for sale to customers." (NIH, 1998, p. 4). The NIH supported the development of a "Uniform Biological Materials Transfer Agreement" (UBMTA) for materials exchanges among academic researchers in 1995, although universities have been slow to adopt it. The National Research Council's 2010 report on "Managing University Intellectual Property in the Public Interest" recommended that universities use either the UBMTA or a Simplified Letter of Agreement (SLA) for materials transfers (National Research Council, 2010, p. 9).

[5] Consistent with this characterization, more than 73% of respondents to the survey by Lei et al. (2009) reported using MTAs for more than 60% of the research tools that they obtained from industry in 2004, while only 35% of respondents did so when they obtained them from academic researchers.

[6] "BRCs balance IP rights against the need for access through materials transfer agreements (MTAs), which offer nonexclusive licensing rights to BRC users. By enhancing the effectiveness of the market for the exchange of licensed materials, BRCs have increasingly come to serve as key knowledge brokers for researchers throughout the life sciences." (Stern, 2004, p. 82).

[7] "This commercialization of research materials may actually increase access by creating market-based institutions for distributing them rather than relying on gift exchange among researchers. Several university scientists noted that the demand for important research agents can easily become overwhelming, and licensing these to a commercial firm was seen as a way of increasing, rather than limiting, access for the research community." (Walsh et al., 2003, p. 322).

research activities. But researchers did report that when their requests for research materials were not fulfilled, their ability to pursue research was constrained. Denial of requests for materials were especially problematic for researchers working with "signaling proteins," a field characterized by high levels of academic patenting and considerable promise for applications in the pharmaceuticals industry.

Walsh et al. (2007) provide no information on the extent to which MTAs covered transfers of research materials associated with patented research results, although as noted, they did find that failures to provide research materials were more pronounced for researchers in signaling proteins. The survey results also reported that more than one-quarter of the MTAs that were negotiated took more than one month to finalize.[8] Our discussions with licensing professionals in the UC system suggest that MTAs with private-sector researchers often take much longer to negotiate and are more likely to involve more complex restrictions. If access by other researchers to patented research results is limited by restrictive MTAs and/or the use of these research materials is constrained by the terms of an MTA, citations to scientific papers linked to these MTAs might decline because of limited access by other researchers to research tools and related materials. Moreover, if the existence of MTAs covering research related to publications associated with invention disclosures is correlated with the likelihood that these disclosures are patented, it is possible that some portion of the "chilling effect" on citations highlighted in both Murray and Stern (2007) and Sampat (2005) may reflect the effects of MTAs, rather than primarily or solely those of patents. With the exception of one preliminary examination of this possibility (Mowery and Ziedonis, 2007), the correlation between patents and MTAs has not been examined, and the effects of MTAs on scientific communication have not been addressed by scholars. Alternatively, MTAs, especially when they involve biological resource centers, may serve to facilitate the exchange of research materials among scholars. We therefore approach the analysis of the effects of MTAs on scientific communication without strong priors as to the likely sign or magnitude of any such effect.

Unlike patents, MTAs are not published once finalized. The "news" effect of a patent issuing on a scientific discovery that is also published (and highlighted in Murray and Stern (2007) as an important justification for their analytic approach) thus may be more attenuated for the first MTA concluded that covers materials related to a patented scientific discovery that has been published. Our statistical analysis assumes that in the absence of formal publication of an MTA, "news" of its existence reaches other researchers through informal communications.

---

[8] Data from the University of California, Davis technology transfer office indicate that MTAs between UC Davis researchers and private-sector entities are more likely than those among academic researchers to require more than 50 days to finalize. This is true of both incoming MTAs, when Davis researchers request materials from private-sector researchers, and 'outgoing MTAs,' where UC Davis researchers provide materials to industrial researchers. These data exclude MTAs that are never finalized, which in some cases may represent a failure to come to acceptable terms for the material exchange. Negotiations for materials transfers to or from private-sector laboratories are also less likely to be finalized.

It is also important to note that our data (discussed below in greater detail) enable us to measure the "effects" of only those MTAs whose existence is recorded by the campus technology licensing offices within the UC system. In other words, we omit a significant number of MTAs that are negotiated informally among researchers without the approval of the campus licensing offices. Perhaps more importantly, we do not observe the effects of the failure of researchers or their licensing offices to negotiate an MTA—we observe only the contracts that come into being, not those that never appear.

III. The effect of licenses on scientific communication

Licenses on university-patented discoveries affect commercialization efforts by companies, and therefore may influence corporate R&D in related areas. Unlike patents, licenses are not published or otherwise subject to mandatory disclosure, and in many cases the identity of licensees is treated by university technology transfer offices as confidential. Why might licenses affect the behavior of academic researchers in formulating their research agenda?

This issue has received little empirical attention from scholars. A recent analysis of patent citations to University of California patents that were licensed (Drivas, et al., 2011) found that citations to these patents by non-licensees increased after exclusive licenses (either by geographic area or field of use) were issued on these patents. Drivas et al., interpret the increase in citations as a reaction by other patent applicants to the demonstration of potential commercial value signaled by the negotiation of a license for the patent. A similar signaling effect could increase citations to scientific articles associated with the licensed patent. In this case, the issue of a license "demonstrates" that a particular area of research has potential scientific or commercial value, leading other investigators to pursue work in closely related fields. Indeed, it is plausible (as various scholars have speculated, with limited evidence thus far) that in the wake of the Bayh-Dole Act, academic researchers may choose research areas based on their potential for private profitability. Regardless of whether a license "signal" operates through perceptions among researchers of scientific or commercial potential, this argument predicts an increase in citations to a paper linked to a licensed patent following the negotiation of the license.

Equally plausible arguments, however, can be developed to predict a chilling effect of licensing on scientific communication. Reactions by university technology licensing offices and/or their licensees to any evidence of patent infringement (even for research purposes, inasmuch as the research exemption from such infringement suits remains unclear) may be swifter and stronger in the case of patents that are licensed. And nonexclusive licenses for academic invention disclosures have many of the features of restrictive MTAs, including provisions on reach-through royalties and limitations on the disposition of any intellectual property resulting from research using the licensed disclosure. Moreover, the negotiation of such a nonexclusive license may take considerable time, delaying access to the materials or tools embodied in the disclosure.

We thus are agnostic on the likely sign of any effect of licenses on scientific communication associated with publications linked to licensed academic patents. Indeed, both effects may be present for papers in various fields of research, and we hope that our results shed light on the magnitude of any offsetting effects.

IV. Methodology

A. Matching invention disclosures, patents, and scientific publications

1. Constructing patent-paper "pairs"

Our empirical analysis adopts a methodology that is broadly similar to that of Murray and Stern (2007). We measure "scientific communication" as the number of times a published scientific paper is cited by subsequent articles published in scientific journals (a 'count'). Our sample includes scientific publications where the researcher disclosed the invention to the university, and where a patent was ultimately issued. We also observe MTAs associated with these patented invention disclosures, which enables us to examine the "before and after" effects of MTAs on publication citations. Data on licenses associated with these disclosures allows a similar analysis of the effects of licenses on publication citations.

Our citations-based analysis provides that we match scientific publications to any related patents, MTAs, and licenses. The connections between patents, MTAs and licenses are easily observable in data collected by the University of California on systemwide "invention disclosures," i.e., the declaration by the university researcher of a potentially commercializable advance. Connecting patents, MTAs and licenses to scientific publications is more involved.

In their paper, Murray and Stern (2007) matched patents to articles published in the journal in *Nature Biotechnology* by reading both patents and the academic articles and using the expert judgment of the reader to link them. Our methodology instead employs 'inventor-based matching' to link patents to scientific papers.

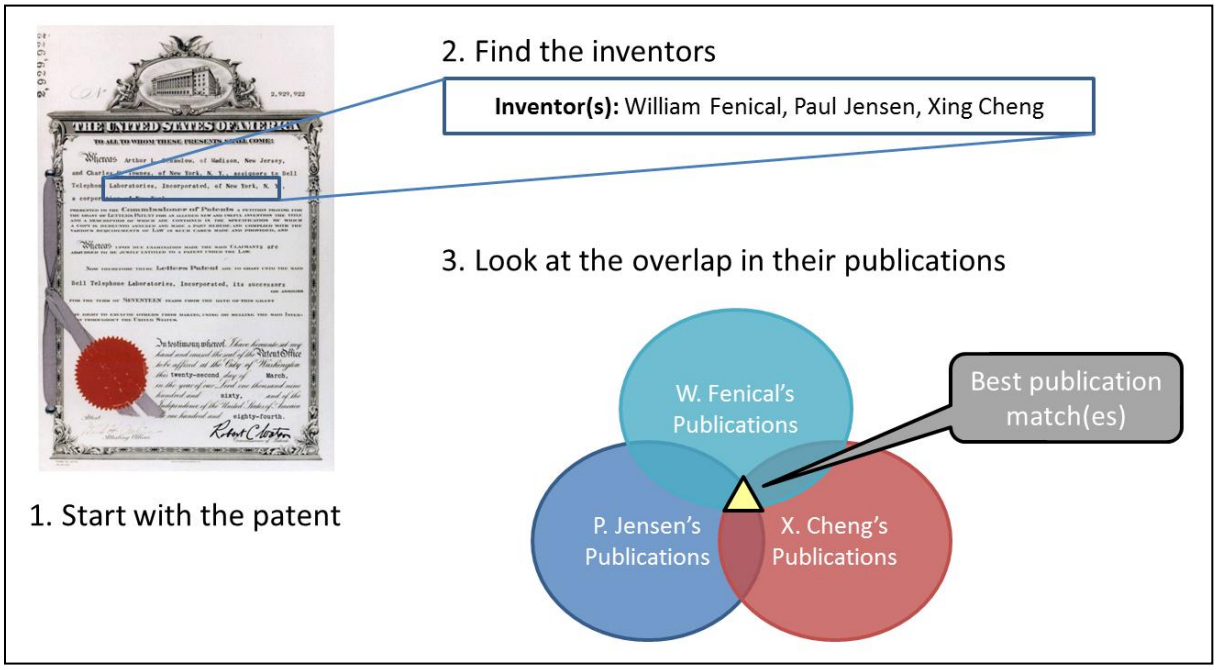
'Inventor-based matching' is based on two assumptions. First, the inventors listed on patent are likely to be the authors listed on related publications. Second, the patent *application* date is likely to occur near the publication date of the academic article.[9] Derived from these principles, we can construct a maximum-likelihood estimator for the publication(s) that best matches a particular patent. We do this by locating the names of the patent inventors, and then identifying all publications authored by these

[9] In contrast to fields such as economics, the time between submission and publication in biomedical research is often no more than a few months (Murray and Stern (2007))

inventors in the year of the patent application or in the two years before or afterwards.[10] We then determine the instances in which the inventors' publications overlap. Those publications with the greatest overlap are chosen as matches.

Thus, if a patent has three inventors, we would extract three publication sets (one for each inventor) and look for publications that are common to all three inventors. Figure 1 schematically outlines this process for the three-inventor case.

Figure 1



This approach can select multiple publications as 'best' matches, in contrast to Murray and Stern, who match one publication to one patent. We can illustrate the value of this generalization with an example from our dataset: a patent on adhesives inspired by the design of gecko feet produces the following matches:

- Adhesive force of a single gecko foot-hair (*Nature*)
- Evidence for van der Waals adhesion in gecko setae (*Proceedings of the National Academy of Sciences*)

Clearly both of these are related to the patent, and thus being able to compare the impact of patent issuance (or other events) on the forward citations of both publications provides additional statistical power.

[10] This ± 2 year period reflects the relative timing of publications and patents observed in the Murray and Stern matching and our replication of it (described below).

Using this matching technique does not restrict us to only instances where all the inventors are listed as authors on the publication - for example, if in the case above a lab technician had also been included on the patent, but was not listed on any of the academic publications. Under these circumstances the algorithm will choose the publication(s) with the maximum overlap possible, thus it would choose a publication listing three of the four inventors since there are no four-out-of-four-inventor matches.

This 'best-available' property is a general property of maximum-likelihood estimators, of which this is one. To discuss this further we make explicit the connection between the algorithm and this class of estimators. First we assume that a publication and a patent are more likely to be a match if they share an author, i.e.:

$$p\left(match_{pub_i\,\&\,patent_j}\middle|author_k \in \left(authors_{pub_i} \cap inventors_{patent_j}\right)\right) \geq p\left(match_{pub_i\,\&\,patent_j}\right)$$

Here $authors_{pub_i}$ and $inventors_{patent_j}$ are the sets of authors for publication i and the inventors for patent j respectively. Given this, pub_m is a 'match' for $patent_j$ if

$$m \in \underset{i}{\operatorname{argmax}} \prod_{i,k} p\left(match_{pub_i\,\&\,patent_j}\middle|author_k \in \left(authors_{pub_i} \cap inventors_{patent_j}\right)\right)$$

As we've highlighted above, an implication of this method is that a single patent can be associated with more than one publication. This will occur precisely when multiple publications share the same level of overlap between the inventors, and when no publications have a greater overlap.

2. Limits of Maximum Likelihood Estimators

As with all maximum-likelihood estimators, just because an estimate is 'best' doesn't mean it is correct.[11] In general, accurate matches are less likely when the precision of the estimate is low. For example, our matching algorithm would produce incorrect matches if, on a four-inventor patent, we identified only publications that listed a single inventor as an author. In this case, the algorithm would theoretically identify as matches *all* publications by *all* of the inventors in the relevant five-year window.[12] To avoid such errors we restrict our matches to only high-precision estimates. We do this by limiting our sample to instances where three or more of the inventors are listed on the publication. The logic behind this criterion is illustrated in Figure 2, which portrays the number of papers matched to each patent in our dataset and lists the number of names common to both the patent and the published paper.[13]

---

[11] See, for example, Casella and Berger (2002) for a discussion of maximum likelihood estimators

[12] The five-year window includes the year of the patent application, the two years prior and the two years afterwards.

[13] The data in Figure 4 include only papers and patents linked by at least two inventor or author names.

Figure 2

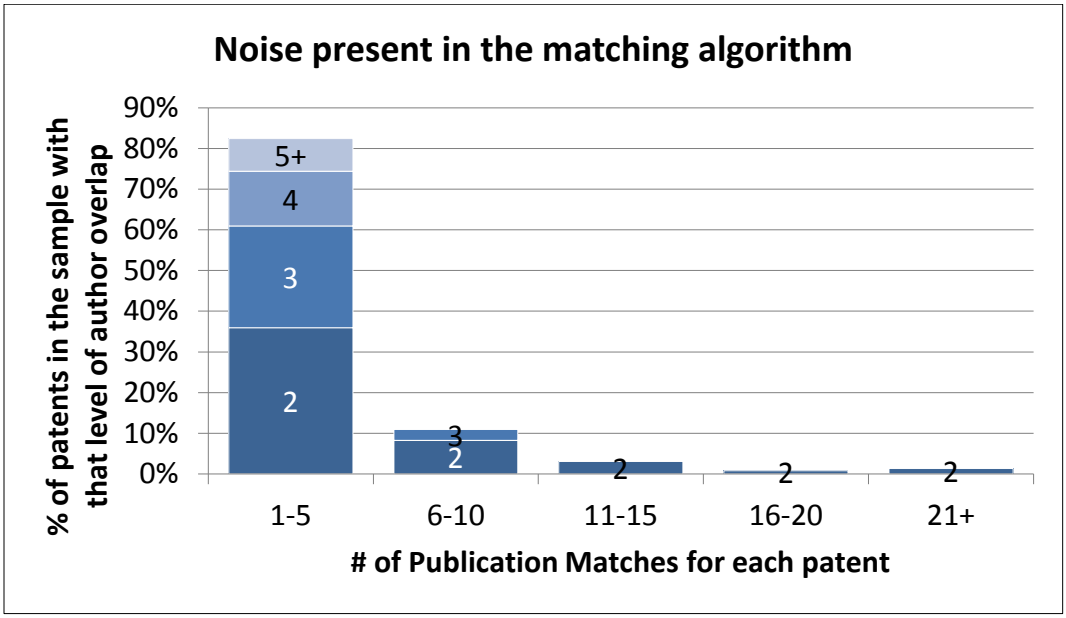


Figure 2 shows that 82% of the patents in our sample are linked by the 'inventor-based matching' algorithm with 1–5 papers, while the remaining 18% of our patents are associated with six or more publications. One likely cause for the large number of papers associated with each of the patents in this 18% is common scientist names (e.g., "Professor J. Smith"). Figure 2 also demonstrates that the majority of these patents with more than 5 publication matches have only two names that are common to both the inventor list and the author list. We therefore exclude all the patents marked '2' in Figure 2, noting that we have previously excluded patents and papers with only one name in common.

We argue in the next section that this restriction reduces the false matches (hereafter "noise") in our dataset, and that sample characteristics are consistent with this interpretation. It is tempting to then argue to exclude '3' author overlap papers for the same reason. We resist this temptation because of the effect on our sample size, but note that it would be an interesting extension for someone a larger dataset.

3. Statistical Implications of Restricting our Dataset

The statistical implications of dropping publications with an author overlap of '2' is to restrict the sample to a higher expected probability of a match[14], that is, to exclude publications whose expected probability of a match is 'too low'.

[14] This follows directly from the definition of the maximum-likelihood estimator above

Table 1: Sample Summary Statistics

Samples		
Inventor Overlap	3	4
Sample Size		
Publications (000)	1.7	0.6
Patents (000)	0.7	0.3
Publications with MTAs	261	79
Publications / Patent	2.4	1.8
Observations in Life Sciences	49%	44%
Sample Statistics[#]		
Citations per year	11.4	16.2
	(26.1)	(36.3)
Average Impact Factor	8.7	11.0
	(8.4)	(9.8)
Publication year	2000.7	2000.4
	(2.6)	(2.8)
Publication Age	3.2	3.3
	(2.6)	(2.7)
Age at MTA issuance	2.4	2.6
	(2.7)	(2.6)
Age at Patent issuance	3.5	3.5
	(2.0)	(2.0)

[#] values in the parentheses is 1 standard deviation

We examine the effect of restricting our sample to higher levels of inventor-overlap by comparing the sample statistics of 3-inventor overlap with 4-inventor overlap, as shown in Table 1. The average year of publication, age of publication when the citations are observed, and the proportion of the papers in the Life Sciences (principally Biology, Biochemistry, and Medicine) are relatively stable across the samples. This suggests little or no introduction of bias along these dimensions.

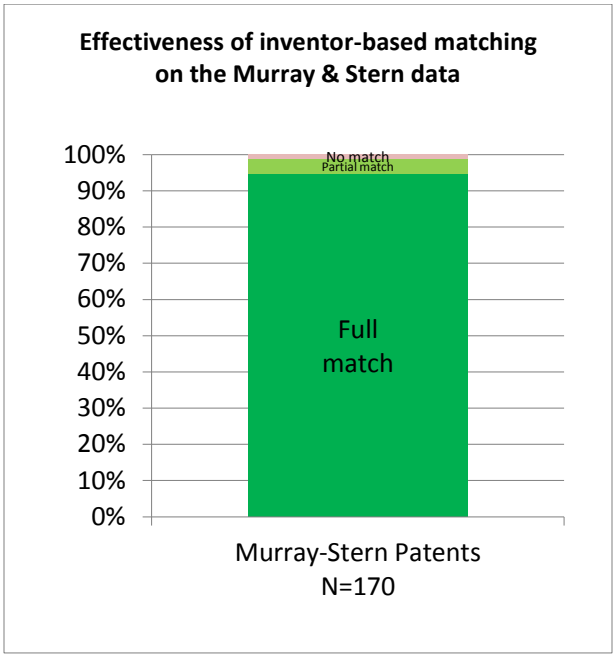
In contrast, two measures of publication quality: the number of citations per year and the average impact factor of the publication's journal do rise in samples with less noise. This is consistent with the observation that patented publications are of higher quality (i.e. receive more citations) than unpatented ones. Since correct matches are patented publications, but incorrect matches are random additions from the general pool of publications, we expect the removal of incorrect matches to produce an increase in average publication quality. This is consistent with our argument that higher-overlap specifications

reduce noise in the sample[15], although certainly not definitive. Summary statistics for our Life Sciences sub-sample (biology, biochemistry and medicine) show these same trends.

4. Testing the quality of the patent-paper matching

Having established the methodology of inventor-based matching, we turn to assessing its validity. As mentioned earlier, Murray and Stern's patent-paper pair sample was developed through hand-matching publications and patents based on the scientific content of each. As such, those matches provide a useful benchmark against which to measure the performance of inventor-based matching, and the gracious cooperation of Murray and Stern enabled us to undertake this comparison.[16] As Figure 3 shows, of the 170 patent-publication matches found by Murray and Stern, the new method identified the identical 'best' publication match as they did for 95% of their sample. In an additional 4% of these instances, the Murray-Stern match was found, but another publication was deemed a better match. In all cases these 'better matches' were in journals other than *Nature Biotechnology*, and thus Murray and Stern would not have considered them in their matching process.

Figure 3



[15] Other things equal, it is plausible that publications associated with patents will tend to be of higher quality than an average unpatented publication. So, if tightening the 'overlap' restriction results in a sample with fewer unpatented publications ('noise'), then it should have a higher average citation rate, which is what we observe.

[16] We are grateful to Professors Fiona Murray and Scott Stern for sharing their dataset with us.

This benchmarking exercise demonstrates the ability of our inventor-based matching technique to identify the best publication matches for patents. Having established its effectiveness, we note several other advantages of this method over that employed by Murray and Stern (2007):

1. It does not impose a simple one-to-one relationship between patents and publications;
2. It enables us to analyze many scientific fields without requiring a domain expert in each;
3. It is a transparent and reproducible matching procedure that does not rely on the gifted intuition of a small group of researchers;
4. It allows the matching process to be automated, making much larger sample sizes feasible.

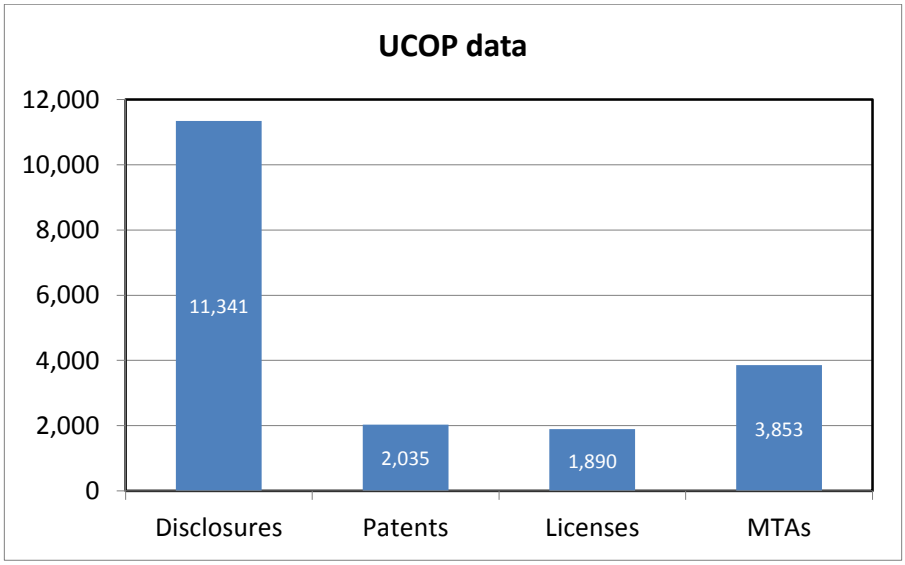
V. Data

A. Sources

We draw on two principal sources of data for our empirical analysis. The first, the 'IP data', is an extract from the technology disclosure database maintained by the Technology Transfer Office within the University of California Office of the President (UCOP). UCOP monitors, and in some cases manages, invention disclosures, patent applications, and licensing transactions for all campuses of the University of California (in this period, nine campuses, which included five medical schools).

Our database extract lists all inventions reported by University of California faculty from 1997 to 2007. Each disclosure is linked to associated patents, MTAs and licenses. Figure 4 presents a summary of the contents of the database.

Figure 4



Note that only a small subset of technology disclosures is patented, and it is likely that patented disclosures are of higher-than-average commercial quality and perhaps also higher-than-average academic

quality. Universities' patenting propensity also varies among fields of academic research — since the 1980s, patenting and licensing activity at UC have been dominated by biomedical research findings. Figure 4 may give the (incorrect) impression that one-third of disclosures produce MTAs. In fact, MTAs are distributed in a skewed manner, with some disclosures generating many MTAs and many disclosures associated with none. This echoes the finding in Mowery and Ziedonis (2007) that MTAs are disproportionately concentrated in biomedical fields of research, as are licenses.

The second source of data, 'Publications data', comes from *Web of Science*, an internet-based service that tracks the bibliographic information and the citations to and from articles published in 10,000 of the highest-impact journals across 256 disciplines.[17] From *Web of Science* we gathered the title, author names, journal, publication date, and citation information for each scientific paper. The information on 'forward citations', citations from later works to that publication, was extracted through the end of 2009. *Web of Science* also provides a number of well-accepted measures of journal quality, which we included in our analysis. The most prominent of these is the 'impact factor', which measures the average number of times an article in that journal is cited in its first two years.

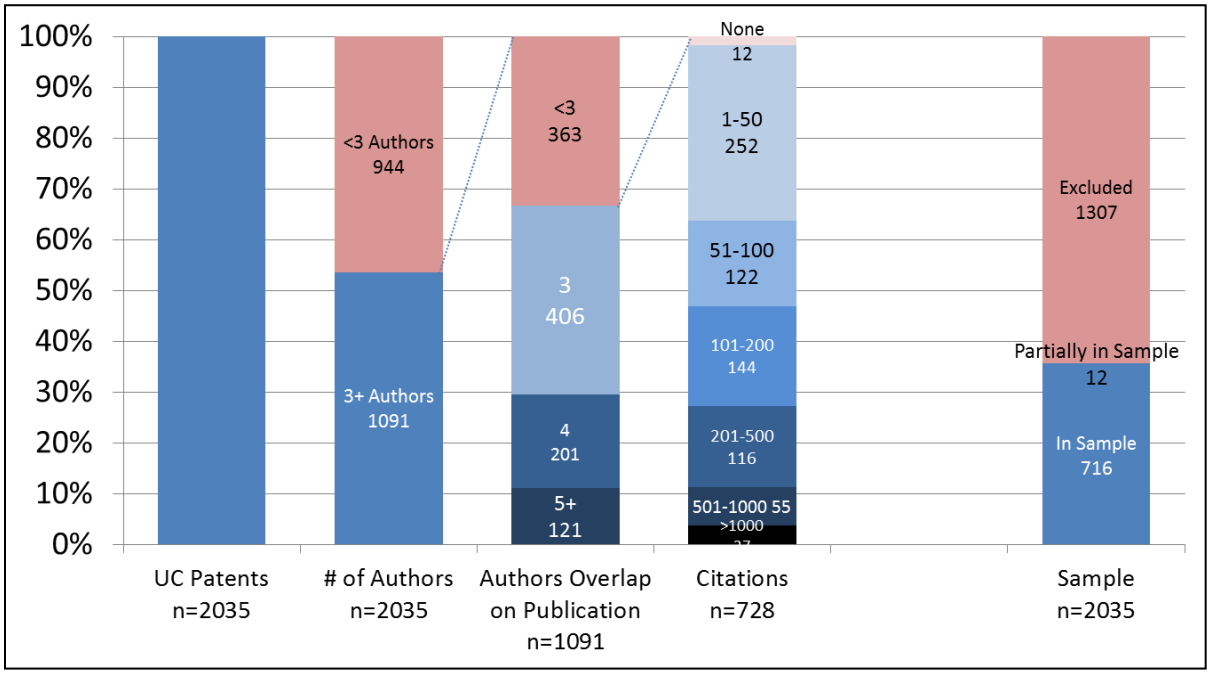
B. Sample

Our study focuses on finding differences in forward citations *within* "patented publications," that is, within the group of scientific publications that are matched to patented technology disclosures. The advantage of restricting our comparison to patented publications is that they are more likely to be similar in quality and other characteristics (including their nearness to commercialization). Although the existence of an MTA or a license for a patent-paper pair may indicate other unobserved differences with patent-paper pairs lacking these features, the scope of this effect should be diminished when we confine our analysis solely to publications linked to issued patents. At the same time, this sample construction means that our estimates measure the effect of MTA issuance on the forward citations of *patented* publications, rather than on publications in general. In future work we plan to investigate this difference between patented and unpatented publications, paralleling Murray and Stern (2007).

Our sample is restricted to patent-publication matches that share at least three authors, which imposes two restrictions on these data. First, a patent must have at least three inventors. Second, the associated publication must list at least three of those inventors as authors. Figure 5 summarizes the impact of these restrictions on the sample:

[17] Web of Science, http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science (downloaded May 2010).

Figure 5: Sample composition



These restrictions exclude 944 patents with one or two inventors and an additional 363 three-or-more-inventor patents where no three inventors were listed as authors on any publication. The fourth bar of Figure 5 also shows the number of citations per patent, i.e., the number of journal citations for all publications that are matched (using the three-name overlap restriction) to that patent. Obviously, publications without any citations will not exhibit any change in citations due to MTA issuance. At the other extreme, a small number of publications receive a large number of citations, creating the risk of outliers that will affect our empirical results. Accordingly, we examine the distribution of residuals for all our analyses to ensure that the results are not driven by a few observations.

VI. "Baseline" regression results

Our baseline analysis adopts an empirical strategy employed by Murray and Stern (2007), applying their technique to the issuance of an MTA or a license rather than a patent. We consider patent-paper pairs that are linked to MTAs or licenses with comparison pairs that are not associated with MTAs or licenses, and compare the number of citations received by publications in the former group in the period prior to the MTA (or license) issuance and contrast this number to those received after issuance. To improve comparability with the Murray and Stern formulation, we restrict our post-issuance period to

the 3 years following the year of the issuance of the MTA or license. Our negative binomial specification is as follows:

$$Citation_{i,t} = f(\varepsilon_{i,t}; \beta_0 + \beta_1 * Age_{t-pubyear} + \beta_2 * Age^2_{t-pubyear} + \beta_3 * Impact\ Factor_i + \beta_4 * Impact\ Factor_i * Age_{t-pubyear} + \beta_5 * Impact\ Factor_i * Age^2_{t-pubyear} + \beta_6 * Patent\ Granted_{i,t} + \beta_7 * NeverIssued_i + \sum_j \beta_j * Journal\ Subject_j + \sum_k \beta_k * Publication\ Year_k + \phi\ IssueYear_{i,t} + \psi\ PostIssue_{i,t} + \omega\ AfterPostIssue_{i,t})$$

Thus, the citations that article i receives in year t are a function of an error term and of a linear combination of an article characteristics and dummy variables corresponding to the timing of the MTA/license issuance (we use the term 'grant' to refer to the granting of a patent and the term 'issuance' to refer to the origination of an MTA or license). The article characteristics include: terms for the publication age (year of observation minus year of publication), including a linear and quadratic form both separately and interacted with journal impact factor, the journal impact factor, fixed effects for the journal subject (e.g. Chemistry) and for the year of journal publication, and finally a dummy variable for whether the related patent has already issued. The terms relating the issuance of the MTA/license are: $IssueYear$, which is 1 in the year that it issues, and 0 otherwise, the $PostIssue$ period, which is 1 in the 3 years after the issuance and 0 otherwise, and $AfterPostIssue$ which is 1 in any periods later than 3 years after issuance. We also include $NeverIssued$, which is 1 if the article never receives an MTA/ license and 0 otherwise. By constructing our dummy variables in this way, our omitted category is $PreIssue$, so our coefficient estimates represent a change versus this default. The coefficient of interest for our analysis is ψ, which reflects the average difference in citations in the three years after issuance as compared to this pre-issue period. We calculate the effect of a license similarly to the above, but with the issue variables recoded to reflect the timing relative to the license.

Our specification differs from Murray and Stern's main specification in that instead of article fixed effects, we explicitly include article characteristics and terms that interact these with publication age. This choice allows greater flexibility in adjusting to changes in citation rates over the life of the publication, but less flexibility in adjusting to between-publication differences that remain after controlling for journal quality. Secondly, we include a dummy variable for whether or not a patent has already been granted, since all our publications are associated with disclosures that are patented and we want to account for the impact of patent issue on future citations to the articles. We present the results from the baseline regressions in Table 2 and Table 3 below.

A. The effects of patents on publication citations

As noted above, our data consist entirely of publications associated with invention disclosures that are patented, in contrast to Murray and Stern, whose sample includes both patent and non-patent-linked publications. We thus are able to observe a 'post-grant' period for the treatment observations but have no equivalent (unpatented) comparison group. The causal influence of patenting cannot be disentangled from age-related effects, therefore, and consequently we do not report estimates for this effect. The $PatentGranted$ variable in the reported results below therefore should not be interpreted as an estimate of the patent effect.

B. The effects of MTAs on publication citations

We report results for the impact of MTA issuance on three samples: the full sample, publications only from the Life Sciences, and publications only from outside the Life Sciences.

Table 2: MTA Effect (baseline specification)

	Full Sample	Life Science	Non Life Science
Patent Granted	1.65 (0.167)***	1.21 (0.140)*	2.04 (0.321)***
MTA Never Issued	0.85 (0.095)	0.874 (0.10)	1.20 (0.288)
MTA Issue Year	0.65 (0.073)***	0.673 (0.075)***	0.82 (0.182)
MTA Post Issue (1-3 years)	0.70 (0.085)***	0.68 (0.101)***	1.07 (0.210)
MTA After Post Issue (4+ years)	0.62 (0.124)**	0.748 (0.224)	0.97 (0.319)
Age	1.59 (0.079)***	1.594 (0.099)***	1.68 (0.111)***
Age^2	0.96 (0.005)***	0.948 (0.007)***	0.96 (0.006)***
Publication Year	1.03 (0.020)	1.053 (0.024)**	1.02 (0.027)
Impact Factor (2006)	1.08 (0.008)***	1.077 (0.010)***	1.08 (0.011)***
Impact Factor (2006) * Age	1.00 (0.003)*	0.999 (0.003)	0.99 (0.004)**
Impact Factor (2006) * Age^2	1.00 (0.000)***	1.001 (0.000)**	1.00 (0.000)***
Journal Subject Dummies	Yes	Yes	Yes
n=	2853	1133	1720

The coefficients reports are incident rate ratios (IRRs), and thus should be interpreted as the *multiplicative* effect on citations. Thus the coefficient on $PostIssue$ of 0.70 for the full sample reflects a 30% decline in the number of citations due to the MTA Issue (which is significant at the 1% level). In the Life Sciences we see a slightly larger effect (0.68 – a 32% decline) that is also significant at the 1% level, whereas the Non-Life Sciences shows a small positive but not significant effect.

C. The effects of Licenses on publication citations

We also report the effect of Licenses on the forward citations of related publications. Our results suggest larger, but statistically insignificant, positive effects in the Full Sample and the Non-Life Sciences, and a mildly negative, but also statistically insignificant, effect in the Life Sciences.

Table 3: License effect (baseline specification)

	Full Sample	Life Science	Non Life Science
Patent Granted	1.57 (0.176)***	1.26 (0.161)*	1.99 (0.328)***
License Never Issued	0.75 (0.092)**	0.54 (0.095)***	0.96 (0.160)
License Issue Year	1.07 (0.097)	0.99 (0.097)	1.12 (0.156)
License Post Issue (1-3 years)	1.17 (0.121)	0.97 (0.103)	1.30 (0.209)
License After Post Issue (4+ years)	1.07 (0.153)	0.69 (0.103)**	1.49 (0.316)*
Age	1.55 (0.087)***	1.60 (0.098)***	1.60 (0.124)***
Age^2	0.96 (0.005)***	0.95 (0.006)***	0.96 (0.007)***
Publication Year	1.01 (0.019)	1.03 (0.023)	1.01 (0.025)
Impact Factor (2006)	1.08 (0.008)***	1.06 (0.010)***	1.08 (0.011)***
Impact Factor (2006) * Age	1.00 (0.003)**	1.00 (0.003)	0.99 (0.004)**
Impact Factor (2006) * Age^2	1.00 (0.000)***	1.00 (0.000)**	1.00 (0.000)***
Journal Subject Dummies	Yes	Yes	Yes
n=	2853	1133	1720

D. Discussion

Perhaps the most notable aspect of the above results is the magnitude of the estimated "effect of MTAs". A decline in the citation rate of 30% for Life Science publications due to the issuance of an

MTA represents a substantial reduction. We suspect that a true effect of this size would evince an outcry from the scientific community, and interviews with TLO officials do not support this finding. The next section discusses sample selection issues that may be responsible for this divergence between our expectations and the estimated effects of licenses and MTAs.

VII. Sample selection issues

A. Constructing the counterfactual using regression 'controls'

To ensure that the control group is indeed a good proxy for what would have experienced by the publications in the treatment group had they not received an MTA or License, we examine differences between the treatment and control groups' observable characteristics, particularly differences in the number of citations that each group received *before* receiving the MTA/License.[18] If these are equal, it suggests good covariate balance, and thus we would expect less risk of bias arising from the observable characteristics. On the other hand, differences could indicate that unobservable differences, for example in the underlying "quality" of the research disclosure, remain between the treatment and control samples. If these quality differences are correlated with the presence of MTAs/Licenses and the citation pattern of the publication, our estimates are likely to suffer from omitted variable bias.

We test this proposition by comparing the number of citations received in the year after publication for both our treatment and control groups. In doing so, we follow the literature by incorporating controls for publication age and year, and for the discipline and impact factor of the journal the article was published. We then compare these using a *permutation test*, which can be thought of as a t-test without the assumption of a normal (or other) distribution.[19] This test reveals that the difference in the mean citations of the groups, 2.75, is so large that the test rejects the hypothesis that they are the same at a 1% significance level. This is strong evidence that, even with the controls we included, a sample selection problem remains. We therefore employ an alternative method to construct our control sample, which we describe in detail below.

B. Constructing the counterfactual using matching

1. Nearest-neighbor matching

To address the sample-selection problem identified above we employ a *nonparametric* approach that weakens the linearity assumptions imposed by the regression 'controls' method, and thus may enable us to achieve improved exchangeability between our control and treatment samples. The specific method

[18] For example, it is plausible that papers that are highly cited by other researchers are more likely to generate requests from these other researchers for related research materials.

[19] See, for example, Wasserman (2005) for a brief introduction.

we use is termed *nearest neighbor* matching. This method searches the set of non-treatment observations to identify the instance that best matches each treatment observation and then places it into a 'Control Group.'[20] This search covers observable characteristics of the observations, and thus, by construction, the control and treatment observations should be similar (we test this below).

Because this matching process begins with the treatment observations, our coefficient should be interpreted as an average treatment effect on the treated (ATT), that is, it is the average effect you would expect from receiving an MTA *on publications that are like the ones that actually do receive MTAs* (in contrast, for example, to what would happen to an average publication). By restricting ourselves to control observations that 'match' the treatment observations, we are necessarily excluding those that are dissimilar, thus decreasing our sample size[21].

We find our matches using the R statistical library "Matching" by Jasjeet Sekhon (Sekhon, 2011), availing ourselves of the functionality that searches for the best matches using a genetic algorithm[22].

2. Criteria

For our nearest neighbor matching we use two sets of variables, those where we require an exact match and those where a nearby match is sufficient. We require an exact match on the following variables:

- Publication Age: number of years since the paper was published
- Journal Subject: academic discipline of the journal (e.g. Medicine)
- Patent Granted Y/N: whether the related patent has been granted
- License Issued Y/N: whether the paper has an associated license at the time of the MTA[23]

These restrictions imply that, for example, when testing for the effects of MTAs on citations, a treatment observation in the life sciences with an issued patent and no license would be compared with a control group observation in the life sciences with an issued patent and no license, and that the comparison would be during the exact year that the treatment observation had received the MTA (e.g. 2nd

[20] For robustness, we also test an n-nearest neighbors approach, where we pick the n best control observations that match the treated observation, and assign each a weight of $\frac{1}{n}$. This produces no notable change from the single nearest neighbor approach.

[21] The effect that this has on the precision of our estimates is ambiguous. Smaller samples will tend to reduce our precision (making the standard errors larger), but there may also be a countervailing impact because the smaller sample will likely be more homogenous, and thus increase precision.

[22] According to Sekhon (2011), "GenMatch dominates the other matching methods in terms of MSE [Mean Squared Error] when assumptions required for EPBR [Equal Percent Bias Reduction] hold and, even more so, when they do not". Previous researchers who have also used this generalization of the propensity score include: Andam, Ferraro, Pfa, Sanchez-Azofeifa, and Robalino 2008; Eggers and Hainmueller 2009; Gilligan and Sergenti 2008; Gordon 2009; Heinrich 2007; Hopkins 2010; Morgan and Harding 2006; Lenz and Ladd 2009; Raessler and Rubin 2005

[23] When considering the effect of a license, this control is changed to whether an MTA has issued.

year after publication). Similarly, tests for the effects of a license would use as control observations those publications of the same age, the same (broadly defined) discipline and the control observations would be chosen based on whether or not they already had issued patents or MTAs at the time of the first license.

Under this system, treatment observations with no equivalent control observation are dropped from the sample. This restriction becomes constraining when the pool of "control" observations becomes small and thus the possibility of finding an exact match becomes less likely—in our case when we consider the MTA and License effects on small subsamples of our data. For those treatment observations that match on the 'exact' characteristics, we then pick the nearest neighbor based on their relative proximity in the following 5 characteristics:

- Journal Impact Factor
- Publication Year
- Citations in $t-2$: citations 2 years before the treatment
- Citations in $t-1$: citations 1 year before the treatment
- Slope of Citation curve between $t-2$ and $t-1$

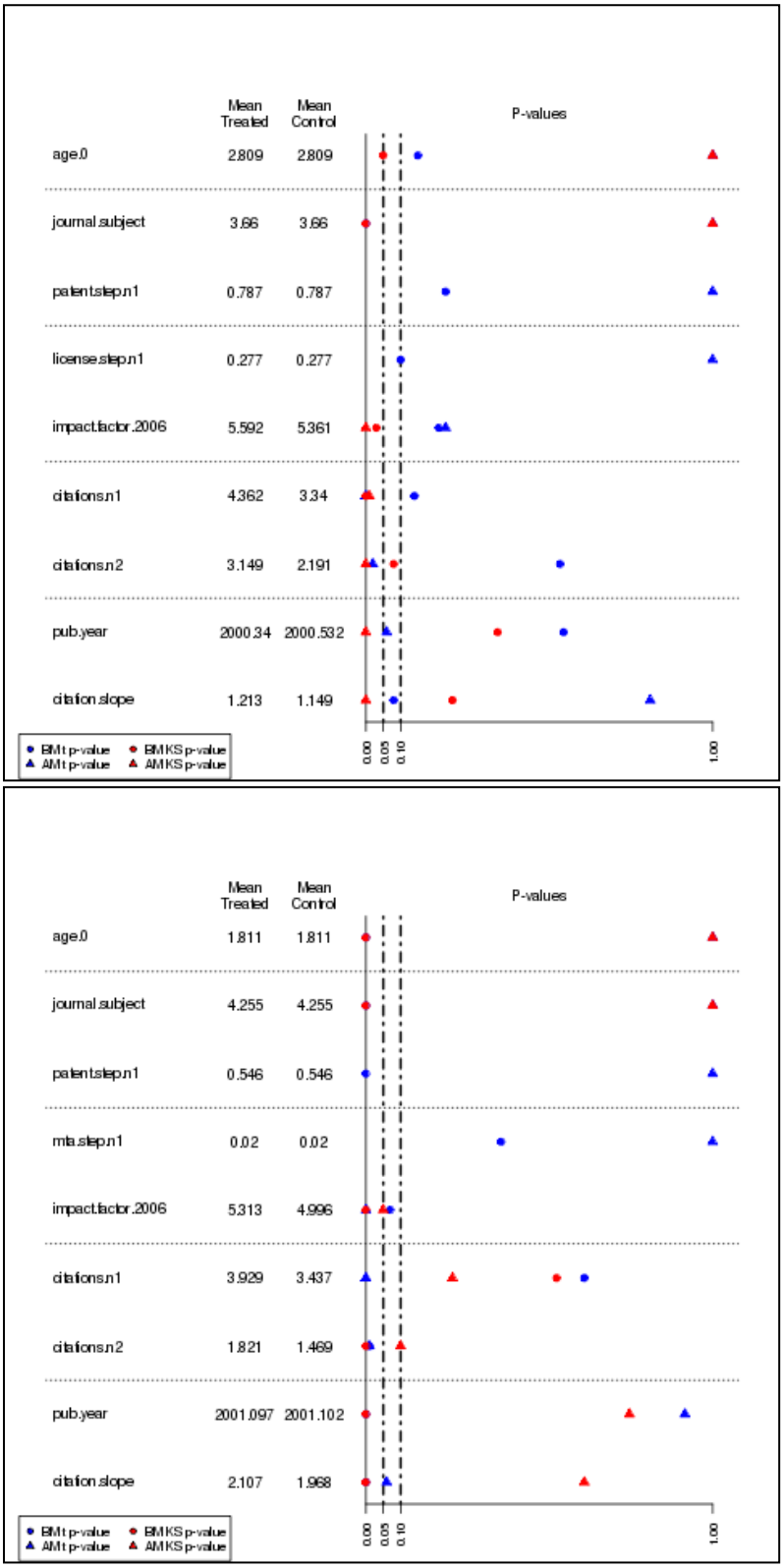
In each of these dimensions we limit the "distance" between each treatment and control observation to 0.5 standard deviations for that variable. Collectively, our restrictions mean that in these 9 important characteristics our control group either matches exactly or within 0.5 standard deviations to the treatment group.[24] Below, we assess the success of this technique.

3. Covariate Balance

To assess the covariate balance between the treatment and control group, we perform two sets of tests: *t-tests* to compare the means of each group and Kolmogorov-Smirnoff tests (*KS tests*), which compare the entire distributions. P-values for the two groups are represented in Figure **6** by blue and red points respectively. For each group we compare the balance beforehand (circles) to the balance afterwards (triangles). The basis for this comparison is the p-value of the test with the null hypothesis that they are the same for the treatment and control groups. Thus, the first four variables have p-values of 1, reflecting that they are exactly matched, and thus have precisely the same distribution and mean between each group. For other variables there may still be differences between the two groups. We can understand the magnitude of these differences by comparing the mean of the treated and control groups. For example, in the MTA sample group we still see differences in citations in $t-2$ (denoted "n2"), and the mean difference between the two groups is approximately 1 citation per year.

[24] The distance limit of an acceptable match thus is the "caliper" of the Matching. Using a caliper helps exclude both observations whose observable covariates would make them outliers and those which would make them inliers, that is observations that are in the 'middle' of the data, but nevertheless lack a comparable control observation (Sekhon, 2011).

Figure 6: Covariate Balance for MTA sample (top) and License sample (bottom)



When we restrict our analysis to part of the full sample, e.g., including only those patents that are in the 'research tools' patent classes (classes 435 and 800), this effectively forces exact covariate balance on this characteristic in addition to the ones listed above. This would mean that each treatment observation is be a research tool and that its matched control observation is also a research tool. This procedure reduces the risk that moving to the "research tools" subsample might introduce bias, but it also means that the composition of different subsamples will differ slightly. In our example, this implies that our estimate gives the additional citation impact of a license *for publications that are research tools*, not for publications ones where the treatment observation is a research tool and its control observation may or may not be. The introduction of additional matching criteria means that coefficients calculated for subsamples are not entirely comparable to those for larger samples.

When differences remain in the characteristics of the treatment and control groups, even after matching, we use *covariate bias adjustment* to control for the differences between the groups. This applies a multivariate linear regression ("regression adjustment") on the post-matching sample (treatment observations plus their matched controls).

To summarize, we obtain our covariate balance using two techniques. First, we use matching to find the nearest neighbor to our treatment observations. This non-parametric technique means that we minimize the (linear, negative binomial, etc.) assumptions that we need to make when constructing the control sample. Secondly, we do a multivariate linear regression (using the same covariates used in matching[25]), to adjust for any remaining differences between the groups. Because we would usually expect these new differences to be smaller, the linearity assumption embedded in least-squares is more plausible. Rubin (1979) discusses the using these (slightly modified) techniques and concludes that "pair-matching coupled with regression adjustment on the matched pairs is a quite effective general plan for controlling the bias due to matching variables, and this combination is clearly superior to regression adjustment" (p.318).

C. Estimator

1. Difference-in-differences

In our baseline regressions above, and in the matched sample analysis below, we estimate our treatment effect size with a difference-in-differences estimator. We compare how the citations to one

[25] For the MTA effect: $\Delta Citations = \beta_0 + \beta_1\, Age + \beta_2\, Journal\ Subject + \beta_3\, Patent\ Granted + \beta_4\, License\ Issued_{YN} + \beta_5 Journal\ Impact\ Factor + \beta_6\, Publication\ Year + \beta_7\, Citations_{t-2} + \beta_8\, Citations_{t-1} + \beta_9\, CitationsSlope_{t-2\ to\ t-1} + \psi\, MTA$. Our coefficient of interest is ψ. For the License Effect, β_4 would be associated with $MTA\ Issued_{YN}$ and ψ with $Licensed$. We report the estimates of ψ in our results.

publication increase (or decrease) following the issuance of a related MTA[26] to the changes in number of citations for a comparable publication that lacks a related MTA. We define this as follows:

$$Treatment\ Effect = (Citations_{t+1} - Citations_{t-1})_{pub\ w/\ MTA} - (Citations_{t+1} - Citations_{t-1})_{pub\ w/o\ MTA}$$

By using a differences-in-differences estimator, we rule out any bias associated with changes that impact the before and after periods simultaneously. For example, in the matched sample case, if a particular scientist gets 5 more citations per year, every year, than this will be added to both $Citations_{t-1}$ and $Citations_{t+1}$ and the impact on our estimate will be zero. In this way, our estimator accomplishes the same effect as author fixed effects. This means that our estimate is robust to these types of *unobservable* differences in addition to the *observable* differences controlled for using matching.

VIII. Results

We examine the impact on forward citations of an issuance of an MTA and of a license issuance. We also examine subsets of the data to see if there are heterogeneous responses within the larger group, for example between publications in the Life Sciences (Biology, Biochemistry, and Medicine) and those outside the Life Sciences (principally Chemistry, Engineering, Computer Science and Physics). We also differentiate between MTAs negotiated with nonprofit research institutions and those negotiated with private-sector researchers. Our analysis of the effects of licenses on citations distinguishes licenses covering patents in the "research tools" classes (USPTO 3-digit patent technology classes 435 and 800) and tests for changes in the "license effect" over time. Ideally we would also consider the impact of MTAs on "research tools" and the changing impact of MTAs over time, but small sample size and non-robust coefficient estimates render such an analysis infeasible[27].

For all results we report the coefficient and statistical significance of the effect from the treatment variable, and omit those for the covariates. We also report Abadie-Imbens standard errors, which account for the uncertainty involved in the matching procedure (Sekhon, 2011). To reduce the impact of outliers on our estimates, we trim the 2.5% top-most and bottom-most citation changes in our data, and then examine our remaining observations using a QQ-plot to ensure that our distribution is approximately normal, and is not driven by outliers. In all cases a single asterisk denotes significant at the 10% level, two asterisks correspond to significance at the 5% level, and three asterisks represent 1% significance.

---

[26] Our analogous test of the effects of licenses on citations to related publications compares the periods before and after the licenses issues.

[27] This may be surprising for those accustomed to thinking of MTAs on research tools as commonplace. We do observe this phenomenon, with 40% of our MTAs issuing on publications related to research tool patents, versus only 20% of licenses issuing on publications related to research tool patents. However that difference is more than overcome by the difference in the far larger number of publications that received licenses as opposed to MTAs. There may be some differences in the difficulty of finding matching control observations.

A. MTA Effect

We first consider how the pattern of forward citations to an article is affected by the issuance of the *first* MTA related to its associated invention disclosure.[28] We focus on the initial MTA for both substantive and technical reasons. Inasmuch as the change in citations occurs due to a change in the sender's behavior about his or her research (for example by initiating a collaboration with an industry or academic partner), the first MTA is likely to signal a change, but subsequent ones would not. There is also a technical reason for focusing on the first MTA. If, rather than considering the initial MTA only, we took into account *all* MTAs, there would be observations with MTAs in consecutive years, making it unclear when the "pretreatment" observations appear on our timeline. Including multiple MTAs for a single disclosure also raises endogeneity concerns, since early MTAs could impact citations that, in turn, impact the demand for later MTAs. There are 216 publications with associated MTAs in our sample. After dropping those with no comparable control observations (lack of common support), and an additional 5% that are trimmed to discard outliers, we retain 47 MTA-linked publications, which combined with their 47 control observations yields 94 observations in our full sample.

1. Effect of an MTA on the Full Sample and on the Life Sciences

In general, we find no significant effect of MTAs on forward citations to their associated scientific publications (Table 4). Our point estimate for the magnitude of the effect is -0.1 citations (which is not statistically different from zero). Using the standard errors from this estimate we can bound the size of this effect with 95% confidence to between -1.0 citations to 0.8 citations. Combined, point estimate and the associated confidence interval suggest that the impact of MTAs on citations is small. Our test of the effects of MTAs on forward citations to scientific publications in the Life Sciences produces similarly small effects that are not significantly different from zero, and which bound this effect between -0.9 citations and 0.7 citations. Our sample of Non-Life Sciences MTAs is too small to enable us to estimate any effects of MTAs.

Figure 7 shows the citation trajectory for both the treatment and control observations in the two years prior to the MTA and the two years afterwards. The Figure provides a visual depiction of the effects of MTAs within the Full Sample and the Life Sciences Sample, the same samples used to estimate the results displayed in Table 4. It shows that citations to publications in both the treatment and control groups increase after MTAs are issued for the treatment group, but that the increase in citations for observations in the treatment group is (non-significantly) smaller. This is consistent with our reported coefficient of -0.1 for the $t-1$ to $t+1$ period.[29]

---

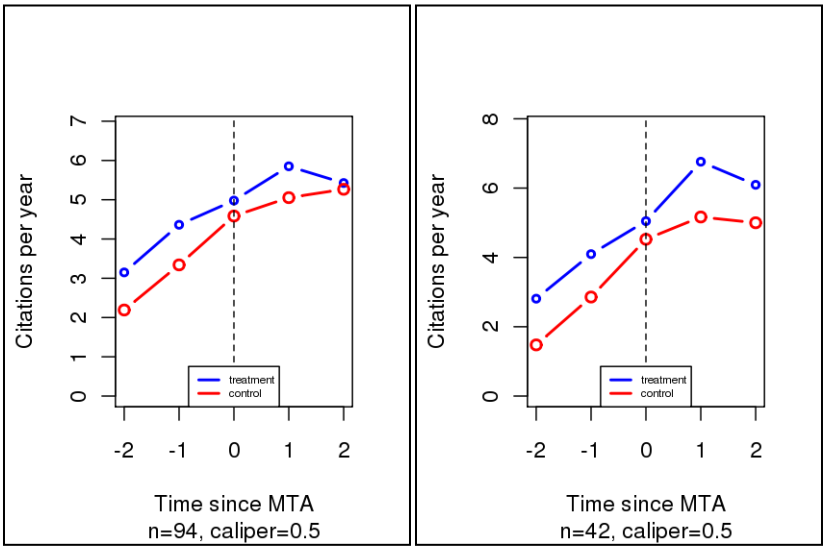
[28] This choice also causes a small reduction in the sample by excluding articles where the first MTA happens before publication.

[29] Prior to a post-matching regression adjustment (see Section VII.B.3), the difference-in-differences estimate is exactly the

Table 4: MTA Effect

	MTA	MTA in Life Sciences
Change in citations (standard error)	-0.1 (0.5)	-0.1 (0.4)
n=	94	42

Figure 7: MTA Effect on (a) Full Sample and on (b) Life Sciences



2. Effect on citations of an MTA with the private sector

We also explore the impact of MTAs with the private sector on citations to scientific publications. This investigation is motivated, in part, by reports from university technology transfer officials that these are the most difficult MTAs to negotiate and that they often contain the most onerous conditions. Consistent with these comments, we observe a negative effect on citations to scientific publications of the negotiation of MTAs with private-sector counterparties of greater statistical significance (Table 3). We observe an overall 1.3 citation decline (7% sig. level) for all MTAs with private sector counterparties, and

---

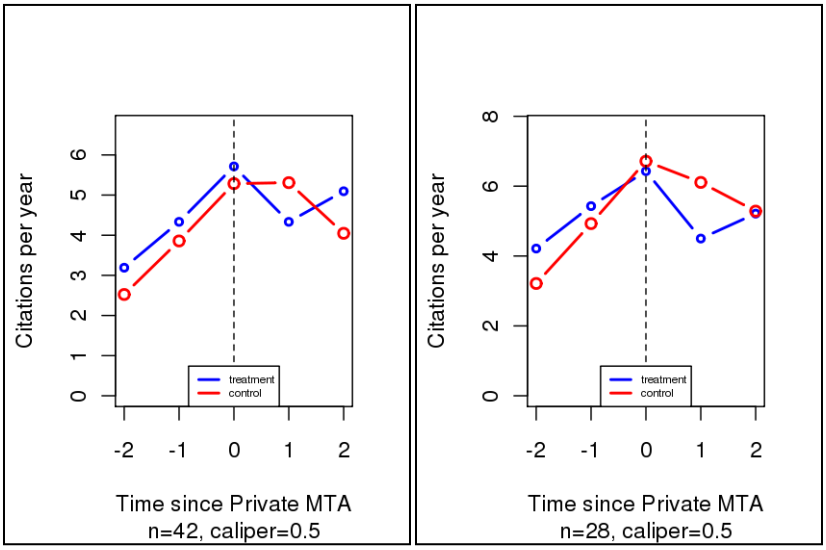
increase in citations to the treatment observations (treatment line) from $t-1$ to $t+1$, minus the increase in citations to the control observations (control line) over this same time period. The post-matching regression adjustment may cause the coefficient to change, however. The extent of the change is precisely the same as that observed in regular regression analysis when a new right-hand side variable is added. Nevertheless, when we see that the coefficient estimate does not adjust (or adjusts only slightly), as is the case between Table 4 and Figure 7, it suggests that the estimates are robust, as they are not heavily affected by the inclusion of the covariates in the post-matching regression adjustment.

a larger 1.9 citation drop (3% sig. level) for Life Science MTAs with the private sector. Table 5 and Figure 8 present these results.

Table 5: Effect of MTA with the private sector

	MTA with the private sector	MTA in Life Sciences with private sector
Change in citations (standard error)	-1.3* (0.7)	-1.9** (0.8)
n=	42	28

Figure 8: Effect of MTA with the private sector for (a) all MTAs (b) MTAs in the Life Sciences



B. License Effect

For the license effect, we again restrict our analysis to the *first* license on a patented disclosure that is associated with a scientific publication. The rationale for this choice is the same as that in our MTA analysis above. There are 1,112 publications associated with licensed patents in our sample. After excluding those observations that lacked comparable control observations (lack of common support), and trimming an additional 5% of the sample to remove outliers, we are left with 392 licensed patent-paper pairs.

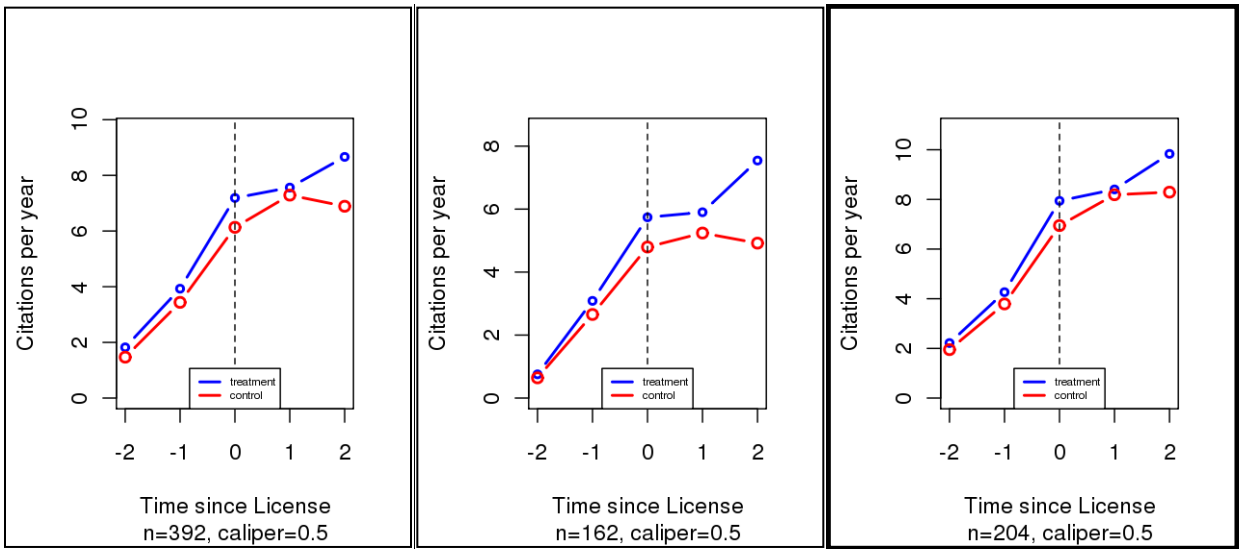
1. License Effect on the Full Sample and within and outside the Life Sciences

As with MTAs, we find that the effect on citations to a scientific publication of a License for the patent associated with that paper is minimal and not statistically different from zero (Table 6). For the Full Sample, our point estimate is -0.3 citations, whereas for Life Sciences it is 0.1 citations and for Non Life-Sciences it is -0.2 citations. Based on this we can bound the effect with 95% confidence to -0.9 to 0.4 citations for the Full Sample, -0.4 to 0.7 citations for Life Sciences, and -1.6 to 1.1 citations for Non-Life Sciences.[30] Table 6 depicts the license effects on forward citations for the Full Sample, the Life Sciences sub-sample, and the Non-Life Sciences sub-sample.

Table 6: License Effect

	License	License in Life Sciences	License in non-Life Sciences
Change in citations (standard error)	-0.3 (0.3)	0.1 (0.3)	-0.2 (0.7)
n=	392	162	204

Figure 9: License Effect (a) Full Sample (b) Life Sciences (c) Non-Life Sciences



[30] Careful readers will notice that the aggregated Full Sample appears to be larger than its two components: Life Science and Non Life Science. This reflects that observations get dropped when they are outside the caliper of 0.5 standard deviations on any covariate. Since the standard deviation changes from the Full Sample to the sub-samples, this may cause some observations to be dropped—this is reflected in the number of observations and explains why a simple average of the two components is insufficient to construct the aggregated number.

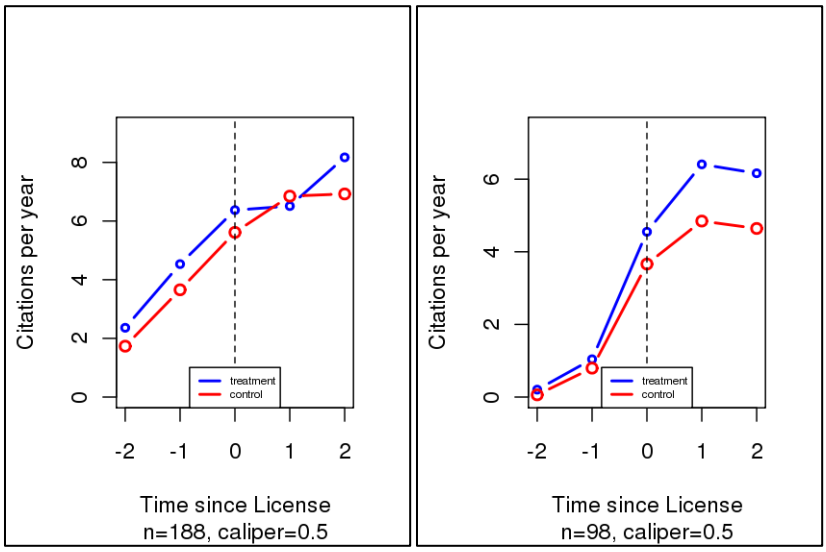
2. Change in the License Effect over time

In line with the Murray and Stern (2008) working paper that suggests adaptation by the scientific community that leads to a reduction in the "chilling effect" of patents on citations to publications linked to patented disclosures by the early 2000s, we explore whether the effect of a license has changed over this period. Murray and Stern test for a behavioral/adaptive shift around 2002, and we report results for the extent of the license effect before and after this break point. We find a statistically significant increase in the license effect from a negative effect in the pre-period to a non-significant positive effect in the post period. This shift is statistically significant at the 1% level.[31] This result is broadly consistent with Murray and Stern's finding that patenting produced lower declines in citations to scientific publications after 2002, although the nature of any change in the behavior of scientific researchers that would produce such a change in the estimated effects of licenses is not clear, and the extent of any "positive" effect in the later period is very modest.

Table 7: License Effect over time

	License Effect pre-2002	License Effect post-2002
Change in citations	-1.0***	0.9
(standard error)	(0.3)	(0.6)
n=	188	98

Figure 10: License Effect (a) pre-2002 (b) post-2002



[31] We check for this using a *t-test*. A superior method would be to compare the two-period version to a single-period version and compare using an *F-test*. We have not yet completed this analysis, but given the strength of the observed effect it is unlikely to impact the results.

3. The effects on citations of Licenses on "Research Tools" patents

Because of the particular importance of research tools as inputs to science, we consider the impact of licensing on patents in patent classes 435 and 800, designated by Murray and Stern (2007) as corresponding to research tools. We report the results of this analysis in Table 8 and Figure 10. Licenses on non-research tool patents have no statistically significant effect on citations to the associated scientific papers, but research tool patent licenses are associated with a statistically significant decline of 3.1 citations per year (1% significance level).

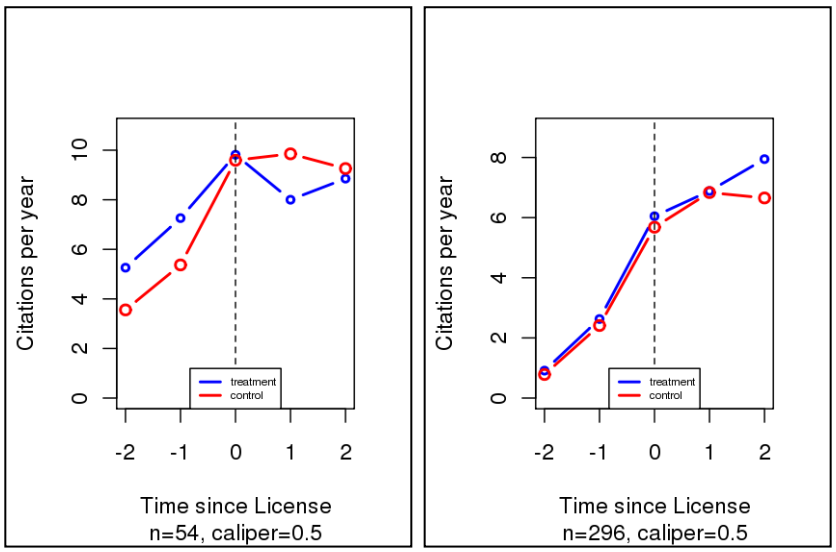
Table 8: License Effect on Research Tools and non-Research Tools

	License Effect on Research Tools	License Effect on non-Research Tools
Change in citations (standard error)	-3.1*** (0.7)	-0.5 (0.4)
n=	54	296

It is worth noting that 96% of the licenses in our research tools patents are exclusive licenses. These results, although estimated on a small sample of patents and publications, suggest that licenses for research tool patents do indeed have a "chilling effect" on scientific communication, perhaps due to an "anti-commons effect" similar to that highlighted by Murray and Stern (2007) for patents.[32]

[32] It is interesting to note that the National Research Council report on university patenting and licensing that was cited earlier recommended that universities "…should try to ensure broad access to research tools." (2010, p. 7).

Figure 11: License Effect on (a) Research Tools (b) Non-Research Tools



C. Discussion

Our results suggest that, in general, MTAs and licenses on scientific publications linked to patented invention disclosures do not hinder scientific communication. Indeed, our results allow us to bound with 95% confidence the average positive or negative impacts to -0.9 to 0.7 citations, and -0.9 to 0.4 citations, respectively. Based on these results, we find little evidence that recent growth in the number of MTAs and licenses covering the results of academic research are constraining scientific communication and the progress of this research.

At the same time, our results suggest that using MTAs and licenses in particular contexts may constrain scientific communication, to the extent that our citation-based measures capture this process. In particular, MTAs with private sector companies are associated with a decline in citations to the scientific papers linked to these MTAs. We also observe a statistically significant and negative effect on citations to scientific publications that appears to result from licenses on their associated patents when these patents are in the "research tools" classes, suggesting some impediments on further research resulting from such licenses. Together, these results suggest that overall, MTAs and licenses have limited consequences for scientific communication. Nonetheless, in some specific contexts, these non-patent instruments may have negative effects on academic science.

On a more positive note, we find limited evidence that the effects of licenses on citations to related scientific publications has switched from negative before 2002 to positive afterwards, perhaps because of adaptation within the research community. Another explanation of these results would be that the 'signaling' effect from licensing has grown in importance, similar to the results of Drivas, et al. (2011) on the effects of licenses on citations by non-licensees to focal patents.

IX. Conclusion

This paper has investigated the effects of Material Transfer Agreements and licenses covering patented invention disclosures linked to scientific papers on citations to those papers. Our results suggest that, in general, MTAs have neither a positive nor a negative impact on forward citations. We do, however, find that MTAs with the private sector, particularly in the Life Sciences, have a significant negative effect on citations. With regard to licensing, we obtain mixed results. Our findings suggest considerable diminution in the significance and magnitude of any negative effects of licensing on citations to related scientific publications after the early 2000s. It is possible that licensing in the later period provides a positive (costly) signal to other researchers about the quality of the research, but these results provide a very limited basis for explaining any such effect. At the same time, however, we see that licenses on research tools show a large and highly significant negative effect on citations.

Our results also suggest that sample selection effects are significant even within carefully constructed treatment and control groups for the analysis of the effects of MTAs and licenses on citations to linked papers. These selection effects may reflect some tendency for researchers to request materials associated with research published in highly cited scientific papers, or growth in citations to papers that results from expansion in researcher requests for materials. Similarly, it is possible that the appearance of a license for a patented disclosure serves as a signal of sorts to other researchers that a particular line of research has particular promise, thereby supporting growth in citations to the scientific papers associated with the licensed patent. Unfortunately, the effects of our sample-matching techniques on sample size and coverage severely constrain our ability to compare the effects of MTAs and licenses among different research fields, which limits our ability to probe these effects.

With the growth in academic research within "Pasteur's Quadrant", fundamental research with potential industrial applications, concerns have risen regarding the effects of patents and other "semi-formal" instruments of intellectual property protection on the flow of scientific knowledge among researchers. As we noted above, the influence of these "nonpatent" instruments, including MTAs and licenses, is unclear a priori. Nevertheless, growth in their number and complexity underscores the importance of additional investigation of their influence on the scientific research enterprise, and we hope to extend this empirical investigation in the near future.

References

Association of University Technology Managers (2003). *AUTM Licensing Survey*.

Blumenthal, D., E.G. Campbell, M.S. Anderson, N. Causino, and K.S. Louis (1997). "Withholding Research Results in Academic Life Science," *Journal of the American Medical Association* 277: 1224-1228.

Campbell, E.G., B.R.Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N.A. Holtzman, and D. Blumenthal (2002). "Data Withholding in Academic Genetics; Evidence from a National Survey," *Journal of the American Medical Association* 287: 473-480.

Casella, George and Roger L. Berger (2002). Statistical Inference: Second Edition (United States: Duxbury)

Drivas, K., Z. Lei, and B.D. Wright (2011). "The Role of Exclusive Licensing in Diffusion of Academic Patented Inventions" Working Paper, U.C. Berkeley.

Eisenberg, R.S. (2001). "Bargaining over the Transfer of Proprietary Research Tools: Is this Market Failing or Emerging?" in R.C. Dreyfuss, D.L. Zimmerman, and H. First, eds., Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society (Oxford: Oxford University Press).

Eisenberg, Rebecca. 2003. Patent Swords and Shields. *Science*, Vol. 299. Downloaded from sciencemag.org in May 2008.

Furman, Jeffrey and Scott Stern. 2006. Climbing atop the shoulders of giants: the impact of institutions on cumulative research. NBER Working Paper 12523.

Heller, Michael and Rebecca Eisenberg. 1998. Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science*, Vol. 280. 698-701.

Katz, M.L., and J.A. Ordover. 1990. "R&D Competition and Cooperation." Brookings Papers on Economic Activity: Microeconomics: 137-192.

Lei, Zhen, Rakhi Juneja and Brian Wright. 2009. Patents versus patenting: implications of intellectual propery protection for biological research. *Nature Biotechnology*. Vol. 27.

Marshall, E. (1997). "Need a Reagent? Just Sign Here," 278 *Science* 212.

McCain, K.W. (1991). "Communication, Competition, and Secrecy: The Production and Dissemination of Research-Related Information in Genetics," *Science, Technology, & Human Values 16*, 491-516.

Mowery, David C. and Arvids A. Ziedonis. 2007. Academic patents and materials transfer agreements: substitutes or complements? *Journal of Technology Transfer*, 32(3): 157-172.

Murray, F., and S. Stern, 2007. "Do formal intellectual property rights hinder the free flow of scientific knowledge? A test of the anti-commons hypothesis," *Journal of Economic Behavior and Organization* 63, 648-687.

Murray, F., and S. Stern (2008). "Learning to Live with Patents: A Dynamic Model of a Knowledge Community's Response to Legal Institutional Change". http://fmurray.scripts.mit.edu/docs/Murray.Stern_LearningtoLivewithPatents.pdf

National Institutes of Health (1998). Report of the Working Group on Research Tools, http://www.nih.gov/news/researchtools/.

National Research Council (2010). *Managing University Intellectual Property in the Public Interest* (Washington, D.C.: National Academies Press).

Rubin, Donald B. (1979). "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*.

Rubinstein, E. (1990). "The Untold Story of HUT78," *Science* 248, 1499-1507.

Sampat, B.N. (2004). "Genomic Patenting by Academic Researchers: Bad for Science?" Working Paper.

Sekhon, Jasjeet S. (2011). "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software*.

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

Stokes, D.E., *Pasteur's Quadrant* (Washington, D.C.: Brookings Institution, 1997).

Streitz, Wendy and Alan Bennett. 2003. Material Transfer Agreements: A University Perspective. *Plant Physiology*, Sept 2003.

Thomson Reuters. Various Pages. *Web of Science*. Downloads in 2010, 2011. www.isiknowledge.com.

Wade, N. (1980). "Hybridomas: A Potent New Biotechnology," *Science* 208, 692-693.

Walsh, J.P., and W. Hong (2003). "Secrecy is Increasing in Step with Competition," letter to the editor, *Nature* 412, 801-802.

Walsh, J.P., A. Arora, and W.M. Cohen (2003). "Effects of Research Tool Patents and Licensing on Biomedical Innovation," in W.M. Cohen and S.A. Merrill, eds., *Patents in the Knowledge Based Economy* (Washington DC: The National Academies Press).

Walsh, J.P., Cho, C., Cohen, W.M., 2005. "The view from the bench: patents, material transfers and biomedical research," *Science* 309, 2002–2003.

Walsh, J.P., Cho, C., Cohen, W.M., 2007. "Where excludability matters: Material versus intellectual property in biomedical science," *Research Policy* 36, 1184-1203.

Wasserman, Larry, *All of Statistics* (New York, NY: Springer, 2005).