BUYER BEHAVIOR IN MARKETS FOR TECHNOLOGY: TECHNOLOGY PROXIMITY BETWEEN FIRM PORTFOLIO AND IN-LICENSED PATENTS

Ayfer Ali

Universidad Carlos III de Madrid

Iain Cockburn

Questrom School of Business, Boston University

1.1 Abstract

Markets for technology promise to increase productivity by better allocating innovative capacity across firms. Research on the demand side of these markets, however, has been limited. In this paper, we use a new dataset of patents available for licensing from a large, innovative academic medical center (AMC) to understand the structure of these markets. Our data includes information on all firms that showed interest in these patents by signing a confidentiality agreement and later decided whether to license or not license the focal technology. Strikingly, we find that of the 285 patents we observe, about 30% of patents available for licensing are never even looked at, and of those that are looked at about 25% are not eventually licensed. Firms with a higher number of own patents and older firms are more likely to take a license. A licensed patent is looked at on average 3.24 times, compared to 2.23 times for patents that have been considered but never licensed.

Because market *safety* issues are ameliorated in this market, we hypothesize that the lack of demand is due partly to the necessity for complementary technologies in the licensing firm. We measure technology complementarity by utilizing widely recognized technology similarity measures which calculate the overlap of International Patent Classes (IPC) between the AMC patent and the firm's own patent portfolio. We find that technological proximity[1] is indeed a determinant of the decision to in-license once a patent has been looked at. Firms are more likely to license technologies that are similar – i.e. "close" – to their own. While this is true when the proximity measure is computed at the broader subclass level of the IPCs, we note that at the more granular, main group level, conditional on subclass-level proximity, greater similarity between the licensee's patents and the AMC patent makes execution of a license agreement less

[1] We use technology proximity and technology similarity interchangeably.

likely. This implies that "close" fit is good but "very close" fit is detrimental for in-licensing. Additionally, we offer improved measures for technology proximity between patent portfolios.

1.2 Introduction

Markets for technologies (MFT), where ideas and early stage technologies are traded, promise substantial allocative efficiencies and opportunities for productivity growth by promoting gains from trade and specialization of innovative labor (Arora and Gambardella, 2010). They are needed when the locus of innovation is outside of the firm best fit to commercialize it. Suppliers of technology can be lone inventors or users uninterested in entrepreneurship, not-for-profit institutions specializing in publicly funded academic research or firms that do not possess the downstream assets to commercialize their technologies in any or all markets (von Hippel, 1976; Bresnahan and Trajtenberg, 1995; Teece, 1986;). On the demand side, potential efficiencies also exist as firms with downstream assets could use their strengths by buying (better) technology from outside instead of (only) relying on their own R&D capabilities (Pisano, 1990).

The potential benefits of markets for technologies can only be realized if they can efficiently provide *stable* matching between each idea for sale and the firm best fit to commercialize it (Gale and Shapley, 1962; Roth, 2008). Market design theorists have pointed out a few characteristics of markets that are needed for such efficiency – *thickness*, *lack of congestion* and *safety*. A market is *thick* if a large proportion of the potential buyers and sellers participate in the market. It is *not congested* if it gives each participant an opportunity to consider multiple transactions. And, finally, it is *safe* for participants when they choose the market over other ways of transacting and reveal their true preferences without engaging in welfare reducing strategic behavior (Roth, 2008).

Gans and Stern (2010) highlight the three main characteristics of ideas that can prevent markets for technologies from operating efficiently – *idea complementarity*, *user reproducibility and value rivalry*. *Idea complementarity* is the notion that ideas are only useful in combination with other complementary ideas. Its existence reduces the number of potential matches to any given buyer or seller and increases the requirements for *market thickness*. *User reproducibility* refers to the fact that once disclosed, ideas can easily be reproduced and the buyer can then become a seller or not pay for the idea (Arrow, 1962). *Value rivalry* is the fact that value gained by one user may diminish as others also use the idea. U*ser reproducibility* and *value rivalry* can reduce *market safety* by inducing strategic behavior by the participants which would result in overall reduction of welfare (Roth, 2008)

Strategy research related to Markets for Technologies has concerned itself mostly with *market safety* issues that may force firms to choose to not transact in the market or can make them engage in strategic behavior (Arrow 1962; Pisano, 1990; Gans et al, 2008; Anton and Yao, 1994; Arora and Fosfurri, 2003; Teece, 1986; Zeckhauser, 1995) In this study we are able to abstract from market safety issues and concentrate on *idea complementarity* and its significance for market *thickness.*

In our paper we explore a small market for technologies in the context of technology licensing from an Academic Medical Center . We observe not only all concluded licenses but also the population of all firms who showed an interest in our sample of patents by signing a confidentiality agreement, evaluative material transfer agreement or an option to an exclusive license. This allows us to describe the structure of demand in a market for technology, something that has never been accomplished before.

This market is special in that problems of *safety* and *congestion* in markets for technologies are alleviated or non-existent. Our ideas are patented providing a good degree of appropriability and reducing issues of reproducibility by non-licensees. Second, while our seller is interested in generating income its overarching goal in licensing is to see these technologies commercialized and serving the greater good. As a result, it is willing to negotiate with the buyer and price is not the reason why a license is not concluded with a potential buyer. Licensing officer incentives are aligned with the goal of commercialization, not profit maximizing, and significant resources and effort are expended in attempt to commercialize these inventions. Furthermore, the institution is in the business of research and patient care and will not compete with the licensor downstream. As a result, it has no strategic reasons to withhold invention related information from the potential buyer. Additionally, asymmetric information problems, especially with regard to uncertainty regarding the technology quality are attenuated – the inventions come from one of the largest and most respected research institutions in the world.

Given the elimination of many *market safety* and *congestion* issues however, we are still faced with a puzzle: of our sample of 285, approximately half (47%) are never licensed and some 85 (30%) are never even looked at. Of those that are looked at, but not licensed, the first firm to look arrives, on average, 2.75 years after the patent has been filed, or approximately 4-4.5 years after the invention disclosure. Of those that are licensed at least once, first license occurs at 4 years after patent filing or approximately 5.5 years from invention disclosure on average. A patent that has been looked at, but not licensed, gets 2.23 looks, while one that has been licensed has been considered for licensing by 3.24 firms and licensed by 2.02 on average.

In this study we show that even when market safety issues have been substantially alleviated, markets for technologies remain *thin* in the sense that a large number of inventions remain not only unlicensed but also never looked at. This leads us to focus on the importance of *idea complementarity* for the efficient working of these markets. We explore the topic by asking the following research question: "How does technology complementarity affect firm decision to buy a specific idea in markets for technologies?"

We hypothesize that a firm's decision to license a particular invention is dependent on how technologically close its patent portfolio is to the patent under consideration. Using widely accepted measures of technological distance we show that firms license inventions that are close to what they own at the broad level of measurement indicating that idea and asset complementarity are important in their decision making process. However, we also find that controlling for broad level fit, a very close fit at the more granular level of measurement lowers the likelihood of a license due to potentially duplicating in-house efforts.

1.3 Literature Review and Hypothesis Development

1.3.1 Markets for technologies

The volume of trade in markets for technologies has been expanding in recent years. Arora and Gambardella (2010) review recent data from various sources to arrive at a market size of approximately $100 billion globally in 2002 which is about double their earlier estimate of $35-50 billion in the mid-1990s. They also estimate that the market has grown at a higher rate than the average global GDP growth rate in the last two decades (Arora et al. 2001; Arora and Gambardella, 2010, cf. Athreye and Cantwell, 2007; Robbins, 2006). Other survey based studies point to the increasing importance and rate of out and in-licensing by firms (Sheehan et al, 2004; Zuniga and Guellec, 2008; Lichtenthaler and Ernst, 2007; Tsai and Wang, 2009)

There is some evidence, however, that not all technologies supplied get licensed. Using PatVal survey data Gambardella et al. (2007) show that 11% of firm-owned patents in the sample are licensed but another 7% remain unlicensed even when the firm wants to license them. While there is no information on firm effort in the licensing process, patent quality differences explain the firm's willingness to out-license a particular patent but not whether a license actually occurs. This leads the authors to speculate that it is market and organizational inefficiencies that result in such a licensing shortfall. The result is consistent with other findings that firms are unable to find interested parties with whom to even start negotiations in 75% of the cases in which they want to license and are able to conclude licenses for only 4% of the technologies they wish to license. They often cite high search costs for licensees as the reason (Razgaitis, 2004).

1.3.2 Demand in Markets for Technologies

There is little information regarding firms' demand for outside technologies in the literature. The few available studies are mostly based on survey data on firm practices rather than specific licenses, use different definitions of in-licensing and are difficult to generalize by geography or industry. Using data from a survey on low and medium technology firms from Taiwan, Tsai and Wang (2009) find that 95% of the 753 firms in their sample licensed technology from outside. Rate of in-licensing also appears to differ by country. While attitudes towards in-licensing are similar between Japan and the UK, for example, the incidence of in-licensing is higher in Japan where companies also search more for technology to in-license (Pitkethly, 2001).

Some studies imply passivity on the demand side of these markets and show that the party that initiates the licensing contact is often the supplier (Atuhanegima and Patterson, 1993). Ford (1988), however, provides statistics without a source that claim that 66% of technology sellers and 45% of technology buyers report that the buyer is the one that initiates the technology deal. Those who in-license seem to value the technology that they have acquired. In a survey of firms using university technology, Thursby and Thursby (2004) find that more than half of the respondents use university technology in new product development and 23% note that in-licensed patents from universities were crucial in the development of their products.

A large portion of the research on the demand side of markets for technologies has focused on the firm's decision to "make" or "buy" outside technologies and the factors that influence that decision. Pisano (1990) shows that the firm's choice of external or internal sourcing of R&D depends on considerations of market *safety*, specifically concerns of appropriability and future hold up due to small-numbers bargaining. Firms are more likely to acquire external technology to shorten product development times and gain competitive advantage especially in fragmented IP-regimes (Atuahenegima, 1993; Kurokawa, 1997; Cockburn et. al 2010).

Other studies however show that the success of a strategy of external technology acquisition depends on in-house R&D investment indicating that the two are complements rather than substitutes (Cassiman and Veugelers, 2006; Lowe and Taylor, 1998; Tsai and Wang, 2007). Internal R&D is necessary not only to be able to absorb technologies that the firm has decided to acquire but also to monitor the state of the technology outside the firm's boundaries and evaluate potential technology acquisitions (Rosenberg, 1990; Cohen and Levinthal, 1990; Arora and Gambardella, 1994).

With our study we contribute to this literature by describing the structure of demand in a market for technologies. We use a new dataset of patents from an academic medical center and observe all instances when a firm showed an interest in a technology and its decision to conclude or not a license for that technology later. While the supply side studies have focused on the importance of the product and its attributes to understand this market, our demand-side focused study lets us also explore firm characteristics in the licensing decision. Specifically we are interested in the importance of technology complementarity in firm decision making. We are able to look at complementary technological capabilities in the firm in a very concrete way by observing the patents that the firm already owns and their characteristics. This allows us to answer the question: "Does the technology developed inside the firm influence its decision to acquire a specific outside technology, given interest in the technology."

1.3.3 The Importance of Complementary Technologies

The importance of complementary assets in firms' technology acquisition decisions has been explored before (Teece, 1986; Pisano, 1990). Two studies by Killing (1978) and Caves et al. (1983) look at how in-licensed technologies relate to a firm's current products and capabilities. They provide descriptive statistics on the type of technologies that firms in-license using a convenience sample of 34 licensee companies in the UK and Canada with over 80 licenses in 1974. They find that 22 percent of the licenses were concluded to strengthen the firm's existing products and 70 percent complemented their current capabilities. However, they only rely on licensee survey reports rather than a technology proximity measure and their definition of proximity relates to the products and firms' capabilities rather than the firms' existing technologies.

Little is known about the influence of a firm's technology portfolio in acquiring innovation from outside. Related studies have looked at the importance of technological proximity for firms' diversification decisions. Breschi et al. (2003) find that a firm's diversification decision is path dependent and firms expand into related fields. Building on the resource based view of the firm, Silverman (1999) also shows that firms diversify into areas where their existing technological resources are most relevant. Furthermore, in the context of strategic alliances, firms whose technologies are more similar to their alliance partners' prior to the alliance tend to "absorb capabilities" from their partners (Mowery et al., 1996). In fact technological proximity has been used to quantify spillovers (Jaffe, 1987).

In a recent study, Laursen et al (2010) assume that firms license technologies that are close to what they currently hold and show that firms with a more diverse current portfolio of technology, implying higher "monitoring" and "assimilation" capacity, will license technology that is further away from their current in-house expertise. However, while shedding some light on the importance of absorptive capacity for in-licensing, their study uses a control group of firms that do not license at all in the period under study. This could lead to significant selection problems. Firms that never showed an interest in licensing may be different at some unobserved level. Our sample, in that sense, provides a significantly better way of understanding the relationship between firms' own technology and what they eventually license as we observe both firms that license and those that show an interest but later withdraw from the market.

Based on findings above that firms may be more willing to diversify into technologically closely related areas, we propose the following hypothesis:

H1: Firms are more likely to license inventions that are close to their own technological portfolio, *ceteris paribus.*

We expect that a firm is better able to know about available technology in an area that is closely related to its current knowledge base, reducing search costs for outside inventions. Furthermore, once such inventions are identified, it will be less costly for the firm to correctly evaluate it and assimilate such outside technology into its current portfolio (Cohen and Levinthal, 1990; Arora and Gambardella, 1994). The firm's existing technological capabilities will then help it extract the most value from it (Silverman, 1999). In this study we don't witness a firm's search for new technology since we only observe firms in the "evaluation" stage. Additional data, in terms of commercialization outcomes will let us observe the process of "value extraction" from the firm's current resources as well.

More importantly however, technology similarity is necessary because ideas are often only useful with other ideas (Gans and Stern, 2010). Heller and Eisenberg (1998) argue that especially in biomedical research, inventions are so interdependent that when intellectual property rights are held by different entities, commercialization can effectively be blocked in case of coordination failure. Such idea complementarity makes inventions only relevant to a few buyers which further lowers chances of a match in the marketplace. As such, the existence of complementary ideas evidenced by technology similarity will be crucial in a firm's decision to license an invention.

Licensing ideas complementary to the ones that it already owns can greatly benefit a firm that is developing new products. However, we expect that technologies that are very similar to what the firm owns in the sense that they can be substitutes to in-house developed inventions will not be licensed. Let's assume that the quality of the in-house and the in-licensed technology are similar and perfectly observable to the firm. The firm has already incurred significant costs for its version of the invention and expects to receive the full amount of the future revenue stream. If it decided to in-license a very similar technology, however, it would most likely pay future royalties to the licensor. As a result, it would choose not to license.

The difficulty of evaluating early stage technologies and the costly transfer of tacit knowledge associated with outside inventions will further lower the chances of a firm licensing even if the quality of the outside invention was better (Polanyi, 1966; von Hippel, 1994; Agarwal, 2006). Furthermore, it is possible that many firms have incentives that reward company scientists for advancing their own technology to the product stage rather than in-licensed technology. Those same scientists are most likely the ones who are evaluating outside technology as well. Behavioral issues such as the so-called "not-invented-here" syndrome which may cause scientists to evaluate outside inventions as inferior to their own have also been pointed out as potential reasons for preferring in-house technologies (Katz and Allen, 1982). This leads us to our second hypothesis:

H2: Firms are less likely to license inventions that are technologically very close (i.e. potential substitutes) to their own technology portfolio, *ceteris paribus.*

We are able to distinguish between H1 and H2 by using an improved version of a widely accepted measure of technological proximity - the cosine, i.e. the uncentered correlation between the technological classes of a focal patent and the firm patent portfolio (Jaffe, 1986). Instead of USPTO patent classes however we use International Patent Classes that have a nested structure and allow us to measure proximity at different levels of granularity. As suggested by previous scholars, we also improve on proximity measures by using all of the IPC codes assigned to a patent rather than the main IPC code (Benner and Waldfogel, 2008).

1.4 Data

1.4.1 Research Setting

The main dataset for our study comes from the technology licensing office (TLO) of a large Academic Medical Center. It contains 285 patents filed and granted from 1980 to 2008 and the associated 307 agreements -- options, confidentiality agreements or licenses -- signed with interested firms for those patents between 1980 and 2010. These patents are the result of employee research and invention. Each employee or affiliate is required to assign to the AMC all rights to all intellectual property developed while at the institution or with funds administered through the institution.

The invention commercialization process starts with an invention disclosure from which a patent is filed which then is licensed through the TLO. When an employee thinks she has developed an invention worth protecting she files an invention disclosure form with the TLO. The invention is then reviewed by a TLO officer with expertise in her subject area who takes on the case. After further consultations with the inventor and further research with respect to the invention's commercialization potential a decision is made on whether to file a patent or release the invention into the public domain. An outside legal firm is then retained to do a patentability search and do the patent filing. As soon as the patent process has been started, the TLO starts looking for potential licensees who will develop their technology further and bring it to market.

There are a few ways in which potential licensees can learn of the invention and these have changed over the years based on new technologies and TLO learning. Brief, non-confidential descriptions of the invention are sent to potential licensees by the case manager after market and industry research. The same description is put on the TLO website where firms can search for it. Firms can also find out about new research results and inventions through research articles and conference presentations by the inventors, through published patent applications or granted patents and of course through direct contacts with the inventor.

Once a firm decides it is interested in a technology, representatives sign a confidentiality agreement (CDA) which gives them access to the confidential description of the research which often includes the patent application and sometimes the invention disclosure as well.[2] The patent application contains valuable information about the invention and the intellectual property rights (IPR) protection strategy. The signing of a CDA does not involve a fee and does not provide an exclusive right to the technology. In fact CDAs with multiple firms at the same time are common. It does however allow the TLO to know of a firm's interest.

Once a CDA has been signed, the firm may return to explore the technology further and reserve the right to license it by signing an option. Options involve some (albeit minimal) fees and often a requirement that the optionee reimburse unreimbursed past and current patent filing and maintenance costs. Amounts can be negotiated and waived if the firm is cash-restrained which is sometimes true of startups. Options normally last less than a year but can be extended for up to an additional year given the right reasons.

A firm can bypass the option stage and decide to sign a license for an invention. Licenses can be non-exclusive, exclusive in field (e.g. diagnostic uses only, a specific treatment area), exclusive, co-exclusive and end user licenses. Licenses usually, but not always, require the licensing firms to reimburse all or a portion of patent expenses and give them the right to participate in decision making on patent prosecution. Licenses provide revenue to the TLO through a combination of a license issue fee, license maintenance fees, milestones payments, percentage of sublicensing fees and royalties. License terms are quite standard based on the technology type but negotiation and variation are possible. Licenses also require that a firm not shelve the technology it is licensing and in addition to maintenance fees and milestone payments may ask for due diligence reports and other evidence of development efforts.

The TLO works with the licensing firm throughout the life of the patents licensed. An exclusive license can be terminated by the firm for any reason and may be relicensed to another firm for development. Amendments can be signed to change due diligence terms or royalty agreement. Sublicenses may be concluded with firms that will develop the invention further or will sell a product in a different market. The TLO does not routinely terminate licenses but if due diligence milestones are not met has the right to.

It is important to note that the licensor in this process is a non-profit institution that does not have the willingness or the ability to compete downstream with potential licensees. One of the explicitly stated goals of the licensor is to see that the technology serve the greater good by being commercialized. As a result, the licensor is interested in maximizing not only the revenue from a potential licensing deal but also gets utility from seeing the technology brought to market

[2] Note that the American Inventors Protection Act granted the USPTO the right to publish patent applications after 18 months from first filing (priority) date. However, it also gives the right to the applicant to request that the application not be published "but only if the invention has not been and will not be the subject of an application filed in a foreign country that requires publication 18 months after filing (or earlier claimed priority date) or under the Patent Cooperation Treaty" - http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp - accessed on November 23, 2011

and curing human disease or facilitating further research. TLO officers have incentives aligned with those goals.

As is the case with most TLOs, if a conflict occurs between academic and commercial goals, academic goals take precedence.[3] In fact, in all licenses the hospital retains the right to practice the invention for research and educational purposes. When a government funded research is licensed, the government also gets a non-exclusive royalty free (NERF) right to the invention and retains march-in rights in service of the greater good. [4]

What this implies for our dataset is that when a license does not occur, it is almost never due to the parties not being able to reach an acceptable price – i.e. negotiate terms. It is because the potential licensee decided that it was no longer interested in the technology for reasons other than price. This is also seen through qualitative data in the case files - comments by officers about why the potential licensee may not have returned for a license after signing a CDA never list price as the reason. Of course, at the very least, to break even the TLO would like to get reimbursed for incurred patent expenses. The utility of the AMC is subject to the costs of patenting, licensing and infringement and license monitoring.

1.4.2 AMC Data

Our data contains all the solely AMC-owned patents that were filed since 1980, after the Bayh Dole act, and were granted by mid-2008. Patents that are jointly owned with other institutions such as universities or companies were excluded as licensing or development activity by the co-owner is not observed in such cases. Patents that are the result of for-profit sponsored research are similarly excluded because sponsorship by industry almost always results in an automatic exclusive option to all the patent rights and in some cases an automatic license with pre-agreed terms which takes the patent off the market. If the sponsoring company did not desire a license after the invention disclosure was made, then that is a signal to other potential licensees that the new technology may not be of high quality since the sponsor would have better private information than a potential licensee that was not involved. In either situation, these patents would not have been directly comparable to the rest and useful for our analyses. This leaves us with 285 patents available for licensing.

The data also includes all the agreements that were signed with for-profit institutions that have the ability to commercialize an invention protected by a patent. These include confidentiality agreements (CDAs) which indicate that a company has shown interest in a certain patent, material transfer agreements (MTAs) for evaluation of biomaterials or prototypes, options to a license for a certain period of time, end-user licenses, non-exclusive licenses, exclusive licenses, sublicenses and patent assignments.

---

[3] Please see a presentation by an officer from another TLO (MIT) for standard TLO goals, http://web.mit.edu/e-club/www/presentations/tlo.pdf

[4] http://www.uspto.gov/web/offices/pac/mpep/documents/appxl_35_U_S_C_203.htm - accessed November 24, 2011

A unique agreement per company and patent was selected if multiple agreements were signed within 5 years of the first agreement with the same company. For example, if a company signed a confidentiality agreement, then followed up with an option and finally signed an exclusive license for the same patent, only the last was selected. If a company signed a CDA but not a license and then 6 years later signed a license, we included both the original CDA and the license as separate agreements. Similarly, amendments to agreements were not included unless they included additional patents and in that case were only included for the new patent.

For our purposes, agreements were divided into two categories – "deals done" and "deals not done." If a license was signed, the agreement was considered a "deal done" and this included exclusive and nonexclusive licenses and sublicenses. Agreements were classified as "deals not done" if they indicated an interest in the patent through a CDA, MTA or option but no license was concluded. This resulted in 307 agreements and overall 600 patent–agreement pairs since many agreements have multiple patents under them and many patents have been looked at and licensed multiple times.

1.4.3 Firm Data

Each firm's technology profile at the time of agreement signing was compiled using patent data. We matched licensee names to patent assignee names conducting assignee name disambiguation by manually going through more than 450 000 company names and by searching for common misspellings. This is important as company names are not standardized at the USPTO. For example, Microsoft patents can be under Microsoft, Microsoft Inc., Microsoft Inc, Micosoft (misspelling) and direct matching to Microsoft Inc. would exclude multiple patents under the other names. In addition, certain companies patent under subsidiary names. For example Zeneca Plant Science and Zeneca Pharmaceuticals are part of the same company and we assume that patents assigned to Zeneca Plant Science are also available for use to Zeneca Pharmaceuticals and vice versa without a license.

Alternative automated ways for name matching are available and we used the 'soundex' function that assigns a string consisting of a letter and numbers to words based on how they sound. We also tried matching using the SAS 'COMPGED' function which measures the "edit distance" between two strings – i.e. the number of deletions, insertions, or replacements in the characters of a word required to arrive at the observed word.[5] In our case, both were inferior to manual matching as they excluded many relevant observations and included irrelevant ones.

For the purposes of this paper, we defined a company's technology position as a stock of patents filed before the time of agreement signing and did not use a depreciation factor for older

[5] http://support.sas.com/documentation/cdl/en/lefunctionsref/63354/HTML/default/viewer.htm#p1r4l9jwgatggtn1ko81fyjys4s7.htm accessed October 1, 2011.

patents. There are a few studies that show that learning depreciates over time (e.g. Benkard, 2000) and company focus may change and lead to a different technological expertise now from the one many years ago. However, it is not clear how long it takes for such technological expertise to change or expire.

One way to determine how long certain technological expertise is relevant for the company is to look at patent validity – if the patent is still valid, then the firm still has that technological ability. However, data limitations prevent us from finding out what patents are still 'alive'. Computation of patent validity at a point in time is impossible without the availability of patent priority data which determines patent term. We have manually gathered that data for AMC patents but we don't have such data for firm patents. In addition, not all patents are maintained to the end of their term. The assignee needs to pay a certain fee to keep a patent alive 3.5, 7.5 and 11.5 years after it is granted.[6] Such payment data is also not available in an aggregated form. Furthermore, some patent terms are adjusted because it takes the USPTO longer to review them. That information is not available in an aggregated form, either.

One difficulty with determining patent portfolio size and content for the companies that are party to these agreements is that most of them are in the pharmaceutical and biotech industry which have seen many mergers and acquisitions in the last few decades. We could not find data on such M&A activity until 1992 and the post-1992 data is not complete for all of the companies so was not included in this version of the paper.[7] The available data is often difficult to interpret as many companies sell or acquire specific plants or businesses such as the "vaccine business" or their "nutritional business" but it is unclear what patents are licensed or sold off with these divestitures and acquisitions. The only exceptions to the exclusion of M&As from our dataset are the top 10 pharmaceutical companies that have seen multiple large mergers and acquisitions in the 1990s – 2000s. Acquisitions of large companies or mergers between two or more large companies were tracked using the web, mainly through company history pages, and verified through Lexis-Nexis Company Dossier service for M&As after 1992. Examples of these include the acquisition of Hoechst by Aventis, the merger of Glaxo Wellcome with SmithKline Beecham and so on.

It is important to note at this point that four companies and two joint ventures with eight agreements and twelve patent-agreement observations were excluded because the company data was not reliable. Regarding the two JVs, it was not clear what kind of parent company patents and knowledge they had access to. Considering them completely unrelated to the parent companies was probably not correct either. All 4 of the remaining companies are highly diversified, with over 30,000 patents each and two of them have a pharmaceutical or medical

[6] This term can be extended to 4, 8 and 12 years with the payment of a fine in the six months between 3.5th and 4th anniversary.

[7] For post-1992 we were able to find some data through Lexis-Nexis Company Dossier service but the data was not complete, especially for non-US firms

device business but the majority of their business is in other industries such as electronics, manufacturing, household goods, aviation, and finance. One of these companies was so large that it had completed 982 acquisitions and divestitures in the 18 years since 1992. Another one had a very common name and it was not possible to disambiguate the name from the other companies in the patent assignee file with the limited resources we had. By mistakenly including patents of unrelated companies, we would mischaracterize the licensee's technological position. Including these diversified companies would have made them not comparable to the other companies and would have distorted our results. This brought down our sample size to 295 agreements and 588 patent-agreement observations.

1.4.4 Patent data

We use patent data to characterize the technology position of a company and the specific AMC technology (Jaffe, 1987; Silverman, 1999). It has been noted that patents are an incomplete indicator of a firm's stock of knowledge as they don't account for expertise that is not protected by a patent. However, multiple studies have used patent data as a proxy for a company's knowledge base using the argument that measures based on knowledge protected by a patent are highly correlated with uncodified and unpatented knowledge (Silverman, 1999; Patel and Pavitt, 1994; Narin et. al., 1987).

Patent data have an advantage in that they are reliably available since 1963 and contain information that can be used to characterize the specific inventions in various ways. The most important information for our analyses is the technology class which has been assigned to the patent. Unlike many other papers that have relied on the USPTO classification system, we use the International Patent Classification (IPC) codes assigned to our patents by the EPO as found in the PatStat database.[8] We also use all of the patent IPC codes rather than assigning the first one to be the 'main' IPC code. Since the EPO itself does not assign a main IPC code we do not worry about weighting one IPC code more than another (Benner and Weldfogel, 2008).

The IPC codes are different from the USPTO classification codes in multiple ways. Most importantly, each IPC code has five nested levels of detail from broad to detailed – section, class, subclass, main group and subgroup level. USPTO classification provides only a class and a subclass. Furthermore, IPC codes divide the technology spectrum into finer slices - at the IPC subclass level which is comparable in the level of detail to the USPTO class level: there are 640 unique subclasses while the USPTO has only about 400 unique classes (Hall et al. 2001, WIPO website[9]). This nested quality is important for our analyses because it lets us measure technology fit at various levels of detail. Furthermore, with the exception of the finest level of

[8] Pat Stat website at: http://www.epo.org/searching/subscription/raw/product-14-24.html, accessed January 2012

[9] WIPO website, FAQ, http://www.wipo.int/classifications/ipc/en/faq/index.html accessed Sept 9, 2011

measurement – the subgroup level (which we don't use) - patent IPC codes are not laterally nested.[10] USPTO patent classes are often laterally nested at the subclass level.

Another important piece of information that is available on a patent are the citations a patent makes to other patents and the citations future patents make to a focal patent. In addition to providing links to other patents such citations are used to determine whether a patent represents a pioneering or more incremental invention. Unlike citations in academic articles, the citations to previous patents, also referred to as *prior art*, delineate and limit the scope of a patent. If a patent cites a prior patent, it means that it cannot lay a claim to the invention in the previous patent. Normally, the more prior art a patent has, the more incremental it is considered and the more developed the technological area to which it belongs. Conversely, fewer backward citations imply that a patent is pioneering. Similarly, forward citations are used to determine the importance of the patent and are correlated with the value of inventions (Trajtenberg, 1990). If a patent is cited by numerous patents, it indicates that many inventors are building on the original invention and hence the invention is more significant (Hall et al., 2001). Furthermore, while only authors can add citations to papers, patent citations are added to the patent both by the patent filer and the patent examiner.

In our paper we use count of patents cited by the focal patent as one of our explanatory variables. The use of forward citations of the focal patent, however, presents some difficulties. To receive forward citations a patent has to be published and the longer a patent has been published the higher chance it has of being cited (Mehta et al., 2001). Through analysis of the overall US patent data Hall et al. (2001) show, that the average patent receives just one half of its lifetime citations by the tenth anniversary of its publication. About 48% of our hospital patents have been granted since 2000 and as such have likely not received even 50% of their expected lifetime citations (Hall et al., 2001). Our data is right censored - the last patent in our dataset was granted in July 2008 and has had only two years to get cited. Because we can't compare total citation count, we construct a variable that measures 'citations per year.' The cites per year variable makes patents of different ages comparable with regards to their forward citations but at the expense of making a strong assumption that the distribution of forward citations is uniform with regard to time. It would, for example, underestimate the importance of younger patents that receive fewer citations early but become very important later on.

A combination of patent citation and patent technology class information has led to the creation of other, composite measures to describe patents. Two important ones are patent *originality* and patent *generality*. The patent originality measure uses the sum of the squared

[10] What this implies is that two patent IPC codes that appear next to each other are not nested within each other at t any of the levels except the subgroup level, which we do not use. For example A61F 2/04 is a subset of A61F 2/02 at the subgroup level. However, A61F 2 and A61F 3 (at the main group level) are not nested within each other. See at http://www.wipo.int/ipcpub/#refresh=symbol&notion=scheme&version=20120101&symbol=A61F0002940000, accessed January 11, 2012.

shares of cited patents that belong to a certain technology class to create a Herfindahl-Hirschman index (HHI) over all technology classes. The measure is 1 minus that HHI. Since we use multiple technology classes per patent, our measure is modified to include not the share of *patents* in each class but the share of *patent IPC codes* assigned to the cited or citing patents. The fewer different IPC codes the cited patents belong to, the smaller the originality measure and less original the patent is considered to be. For example, if a patent that belongs to organic chemistry cites patents that are in software, organic chemistry and aviation, that would be considered a more original patent than one that cites only patents in organic chemistry (Hall et al., 2001).[11]

The generality measure is created in a similar way – only this time, forward cites are used and the HHI is created over the squared share of each IPC code summed over all citing patent IPC codes and subtracted from one. A focal patent that is prior art to patents in multiple different classes is considered more general as it is a platform for multiple technologies in different fields. An example of a very general patent here could be an invention in molecular biology that has been cited by patents in classes as diverse as plants, drugs, electricity, manufacturing and electronics. Compare that to a less general molecular biology patent that is only cited by other patents in the molecular biology field (Hall et al., 2001). The generality measure has a serious shortcoming because most patents have not received all of their future cites at the time of observations and hence their generality may be underrepresented.

Another composite measure created from patent statistics is the patent scope. It is defined as the number of IPC codes assigned to a patent. The more IPC codes a patent has, the broader its scope is considered to be. It has been shown to determine patent valuations by venture capitalists and patent licensing outcomes (Lerner, 1994; Gambardella et. al., 2007; Decheneaux et al., 2008)

1.4.4.1 Technology Similarity/Proximity Measures

The main independent variables that we construct using patent data are our technology proximity measures between the focal AMC patent(s) and the patent portfolio of the interested firm. For this purpose we use the cosine measure pioneered by Jaffe (1986) but with modifications that utilize all IPC codes instead of the main USPTO patent class and at different levels of detail of the IPC code - subclass and main group.

The cosine measure calculates the angular distance between two vectors that characterize the firm's and the AMC patent's position in a technology space defined by patent classes. For this purpose we create a technology position vector for a firm's portfolio $F_i=(F_{i1}, F_{i2}, F_{i3} \ldots F_{ik})$, where each 'entry' is the share of a firm's patents (in our case IPC codes since one patent may have multiple IPC codes) in a certain technology class k. A technology position vector, F_j is also created for the specific hospital patent under the firm's agreement. The angular distance between the two vectors is then the measure of technology similarity and it ranges between zero and one,

[11] We used the correction in Hall et al., 2001 for the calculation of these measures.

one being a perfect fit and zero being no overlap in technology. It is calculated using the following formula:

$$P_{ij}=F'_iF_j/\ sqrt(F'_iF_i)(F'_jF_j)\ ,\ 0<P_{ij}<1$$

For companies that did not have patents filed before the date of the agreement the cosine measure is not defined. We replace the proximity measure for such observations with zeros indicating no fit and run results with and without replacement. We also provide descriptive statistics and models where we exclude observations for which the cosine measure is not defined.

We modify the cosine measure further by computing a "within section" cosine measure constructed based on the above formula, except we exclude all IPC codes in the firm patent portfolio which do not match the AMC focal patent IPC codes at the section levels. For example, if the AMC patent under consideration has the following two IPC codes A61K 9/12 and C07B 12/07, we only keep the firm IPC codes that are in the A and C sections deleting IPC codes in B, F, G and H. This measure looks at proximity of the closest part of a company's technology portfolio to the focal patent. Again firms that are left with no IPC codes that are in the two sections will get a cosine measure of zero.

Both the broad and the within section cosine measures are calculated at the group and subclass IPC code level. However, in our models we use the regular cosine measure at the broader level and the within section cosine measure at the more detailed, main group, level. Our results do not change in direction or magnitude if we replace the within-section cosine measure at the main group level with the regular cosine measure at the main group level (i.e. without excluding the sections that do not match the AMC patent IPC code sections)

1.4.5 Control Variables

For each patent – agreement pair, we also calculate a technology age measure which is the time in years from the hospital patent's patent priority date to the agreement date. The priority date is the date on which the first patent on a certain invention disclosure was filed with the USPTO. This original patent can then be divided into multiple patents if the USPTO deems that it contains more than one separate invention. Continuation patents can be filed from the original patent and continuation-in-part patents, in particular, can add some new matter. [12] Since the priority date is closest to the invention disclosure date, technology age calculated with the priority date rather than patent filing date is an indication of how mature the technology is. Technology risk is more likely to be resolved for older inventions as inventors or others have developed it further to bring it closer to market.

[12] Note that the priority date is NOT the same date as a provisional patent date. For more information on priority dates, please see http://www.yale.edu/ocr/pfg/guidelines/patent/continuation.html, accessed November 29, 2011

An indicator variable for device was generated by looking through each patent's claims. Claims of an apparatus with human body contact or some sort of an implant were marked as devices. The criterion was whether the invention would have required an approval by the FDA as a device in order to be used in the market. For example, an apparatus for growing cells would not be considered a device but rather a research tool. Devices include stents, artificial joints, catheters, surgical instruments, MRI machines and so on.

To describe the technology further, we used indicator variables for the IPC code sections to which the patent belongs. Since many patents have multiple IPC codes that sometimes belong to different sections, we made them into mutually exclusive categories i.e. patents in section A only, patents in section A and C only and so on. Most of our patents are in section A – "Human Necessities" which includes class A61 – "Health; Life Saving; Amusement" which includes most drugs and medical devices. The next important section in our data is section C – "Chemistry; Metallurgy" which under classes C07 and C12 includes organic chemistry and biochemistry, molecular biology and genetic engineering. Section G – "Physics" includes patents related to imaging, ultrasound and measuring. The omitted category in our models is "C&G only" with 22 observations and 4 additional patents that include IPC codes in B or F section.

We also compute a few firm level variables based on patent measures. Because many of our firms, especially those that appear multiple times in our data undergo mergers and acquisitions, we are not able to include firm fixed effects. Furthermore, even though some firms have multiple agreements, most firms have only one or two. Our main concern in controlling for firm differences was to separate the old pharmaceutical companies whose expertise is mostly in small molecule drugs from the biotechnology companies that are specialized in molecular biology. To be able to do this, we created two variables – R&D age at time of agreement and total number of granted patents applied for before time of agreement. The R&D age is defined as the time from the first filing of a patent by the firm to the time that it signs the specific agreement with the AMC. Unfortunately, our patent data goes only to 1963 so big pharmaceutical firms that have been in existence since the 1800s and licensed a technology in 1990 will be at a similar age with a firm that was created in 1980 and licensed a technology in 2007. Similarly R&D age is not defined for firms with no patents.

Another firm level variable that is better at distinguishing the old pharmaceutical companies from the young biotechnology companies is the number of firm patents at the time of agreement. Since old pharmaceutical companies have undergone multiple mergers, they have a substantially larger number of patents than young biotech firms. In that sense the "number of patents" variable is a proxy for both size and age. Because there is a high correlation between R&D age and number of patents, we only include the number of patents variable in our models. Because we expect non-linear effects as well, since this variable's range is very high, we also use its square in our analyses.

We are interested in how the influence of proximity on licensing varies with age and size – i.e. with the type of the firm. Because we have two proximity measures, interaction effects will be difficult to interpret. For that reason, we separate our firms in three groups based on age – early startups, growth startups and old companies where early startups will have filed their first patent between 0 and 10 years before an agreement was signed with our AMC. Mature startups will have filed their first patent between 10 and 20 years before the agreement and old firms would have filed their first patent more than 20 years before the agreement. Based on size, we split the firms in two groups – those that have fewer than 500 patents and those that have more than 500 patents. Those that have fewer than 500 patents are generally biotech or medical device firms. The other group contains all the small molecule pharmaceutical companies.

1.4.6 Descriptive Statistics

1.4.6.1 Patent Level Descriptive Statistics

Descriptive statistics for the patent level dataset are included in Tables 1.1a and 1.1b. The patents are separated into three different groups – those that were never looked at, those that were looked at but were never licensed and those patents that were licensed at least once. Significance levels of two tailed t-tests of comparisons between the first two groups and the "licensed" group are indicated next to the mean of the variable in the respective group. For example, the stars next to the mean value of the variable "Number of Cited Patents" in the "Never Looked At" group indicate that the difference of the means of the "Number of Cited Patents" between the "Never Looked At" and "Licensed At Least Once" groups is statistically significant.

The cohort variables in Table 1b that group patents into cohorts based on their priority date show an interesting story– 88% of the patents that were looked at but not licensed have a priority date after 1995 compared to only 40% of those that have never been looked at and 25% of those that have been licensed. It is important to point out this can influence how the rest of the variables are distributed across group. This observation is not surprising because we expect that patents that have been only recently filed pose a higher risk for commercialization as the technology has not yet been tested and proven. We expect that the probability of licensing increases up to a certain patent age and then decreases as the remaining patent life becomes too short for high investments to be recovered.

The "lead inventor experience" variable is defined as the number of inventions that the inventor has previously disclosed and patented at this specific technology licensing office (TLO). It is a proxy for the inventor's experience both innovating and navigating the licensing process. Many firms can view an inventors' previous experience as a signal of the quality of the invention and its commercialization potential. Furthermore, a larger number of inventions can imply stronger and broader IP rights if the inventor has worked on similar problems before and his previous inventions are related to the current ones. We note that, as expected, the "lead inventor experience" variable is smaller in the "never looked at" group than in the licensed group and this difference is statistically significant. The patent scope variable, defined as the number of unique

IPC classes on each patent (Lerner, 1994) is also intended to measure breadth of IP rights. In this case we find that it differs significantly between the "never looked at" and "licensed" group.

As expected, forward cites per year and two year forward cites are largest among the licensed patents indicating that those are more important and more inventions build on them. However, the result is only statistically significant for the difference between the "Looked At, Not Licensed" group and the "Licensed" group.

"Number of Cited Patents" is an indicator of how pioneering the technology is. Radically innovative patents would have little or no prior art because if a patent cites a previous patent, it means that it builds on it. We note that while 92% (140 out of 152) patents in the "Licensed" group cite prior patents, only 77% (66 out of 85) do in the "Never Looked At" group. This indicates that the more pioneering a patent is, the less likely it is to be looked at or licensed and the trend holds over all groups. This also implies that university technology is ahead of industry developments.

This implication that firms prefer to license in more established technologies even if that leads to narrower IP rights is confirmed by the average age of patents cited by the AMC patents under the agreements. We note that of those patents that cite at least one prior work those in the "Never Looked At" group cite on average younger patents than those in the other groups and this difference is statistically significant.

Patent originality and generality use the IPC classes to which forward cites or backward cites belong. As such when patents have no prior art or no forward cites as of yet, their generality and originality measures are undefined. As a result there are fewer observations in these categories. For originality, we substitute one which equals fully original for patents that have no prior art. While it does not follow the formula by which originality is calculated, it follows the intuitive understanding of originality.

Our generality measure depends highly on the age of the patent. Patents that are younger may not even have a generality measure that is defined because they don't have citations. Older patents, on the other hand, are expected to be on average more general as they have more citations which will have a higher chance of belonging to multiple unique classes. In fact, in our dataset we see that while only 25% of the whole set of patents are missing a forward citation and a generality measure, only 50% of the 'looked but not licensed category', with the youngest patents on average, have any forward cites. And while we see that as a whole they have the highest mean generality measure, this may reflect the influence of outliers in a smaller sample. We can't replace the generality measure with a zero for patents that have no future citations because it is possible that such citations will be received in the future. As such, we don't use the generality measure in most of our models since it significantly reduces our sample size.

Interestingly, while 38% of the patents in the 'never looked at' group are devices, only 29% and 20% in the 'looked at but not licensed' and 'licensed' categories respectively are devices, indicating that devices are much less likely to be licensed. Licensed patents also have a higher number of agreements associated with them – on average 3.24 compared to only 2.23 for the 'looked at but never licensed' group and this finding is statistically significant.

1.4.6.2 Patent-Agreement Level Descriptive Statistics

The next level of descriptive statistics is at the agreement – patent level. In this dataset there are 588 observations and each observation is a patent agreement pair. Each agreement is either a 'deal done' or 'deal not done' and some agreements contain multiple patents. Since some patents also have multiple agreements on them, a patent may be in the 'deal not done' column with a certain agreement and in the 'deal done' column with a different agreement.

Our most important independent variables are the proximity measures. The cosine measure at the subclass level defines the overall broad-level proximity between a firm and a patent while at the main group level it represents proximity at a finer detail. The main group level cosine measure is more appropriate for "within section" fit. Both measures have a range between 0 and 1 with 1 indicating a perfect similarity and 0 indicating no similarity between the focal patent and the firm patent portfolio. We note that our cosine measure at the subclass level is smaller in the "no deal" group than in the "deal" group, indicating that firms license technologies that are closer to the technology that they own. At the cosine main group level measured within the AMC patents' sections, there seems to be no difference between the two groups.

It is important to note that the cosine measures are not defined for agreements with firms that have no patents as they don't have IPC classes for matching with the AMC patent. In Table 1.2a, we first show the mean of this variable after we replace the cosine measure with a zero indicating no similarity between the patent and firm technology for firms that have no patents. The difference of the means of the cosine subclass level between the "deal" and "no deal" groups is statistically significant in this sample. We then exclude those observations where the cosine is not defined and calculate the means without replacement. The means of the cosines are not significantly different between the groups any more.

Firm R&D age differs significantly between the two groups, with older firms more likely to conclude licenses. In fact, it is firms that are younger than 10 years that drive this result as seen in Table 1.2b. They are likely to look at patents but not license. This is also true for firms that have no patents which could be very young i.e. startups but which could also be firms that are not engaged in R&D such as contract research organizations (CROs). Similarly, firms with more patents are more likely to license. The results are also confirmed by the age and size distribution of firms in the "deal" and "no deal" groups in Table 1.2b. In fact, Table 1.2b shows that a larger share of the "no deal" than of "deal" observations are with firms that have no patents or have fewer than 500 patents. Finally, licensing agreements ("deals") include more hospital patents than "no deal" agreements, indicating that firms prefer to license large portfolios of patents indicating more complete intellectual property rights for related inventions.

In confirmation of our previous result on the age of technology offered for licensing, we find that "deals" are significantly more likely to occur when the technology is older, indicating again that firms prefer more established inventions. They also prefer inventions with higher impact indicated by the difference between means in the two groups along the "citations received” variable. As expected from the patent dataset devices represent a higher percentage of the “deals not done.”

The last set of descriptive statistics is at the agreement level. The difference from the previous table is that we are counting each agreement only once. We started with 307 agreements and excluded 12 firms for a sample size of 295 agreements. Again we report cosine statistics with replacement and no replacement. The cosine statistics at the agreement level are calculated over all the IPC codes over all the patents that the hospital has under the agreement. Also, at the agreement level we do not report a within section cosine, but rather a regular cosine at the group level. The results are in the same direction as the ones at the patent-agreement level but the differences in means between the two groups are not significant any more.

1.5 Results:

1.5.1 Patent Level Models

Because we have some patents that are “never looked at”, some that are “looked at but never licensed” and some that are “licensed at least once”, we run a number of analyses to understand what patent characteristics may influence licensing or interest in an invention in general. We run a multitude of regressions using the variables that we have described above. The results are in Tables 1.4a and 1.4b.

The first models we run, reported in Table 1.4a try to explain what variables influence whether a patent has been looked at by a firm, how many times it has been looked at and how much time has passed before the first look. Each agreement, whether a CDA, MTA, option to license or a license is a “look”. The second set of models, in Table 1.4b has licensing outcomes as the dependent variable – whether a patent has been licensed, how many times it has been licensed and time to first license. The first models in Table 1.4a and 1.4b are logit models where the dependent variable is whether a patent has an agreement on it i.e. has been looked (Table 1.4a) and whether a patent has a license on (Table 1.4b). Interestingly, we see that the variable that measures the number of citations received, an indication of importance, is not statistically significant in explaining whether a patent is looked at or licensed. However, the number of previous cites is positively related to whether a patent is looked at or licensed, indicating that more established technologies are more likely to be successful in markets for technologies. We also note the significance of the lead inventor experience pointing to the importance of quality signals in a market with a lot of product uncertainty

Similar to the descriptive statistics, we see that patents that are devices are less likely to be successful. This is unexpected given the importance of physicians in new medical device developments (Chatterji et al 2008). It could, however be due to the fact that new devices do not always require AMC resources or government funding to develop and the best ones may be patented outside of the AMC technology commercialization process. Alternatively, the successful ones could be developed in close collaboration with industry under sponsored research agreements and may thus be excluded from our dataset.

The results are repeated in the next two models – the Poisson and negative binomial- where the dependent variable is the number of times a patent has been considered for licensing or licensed. Because from Table 1.1a we see that both dependent variables "times looked" at and "times licensed" are over-dispersed with the standard error slightly higher than the mean, we conduct a likelihood ratio test which shows that the negative binomial, rather than the Poisson, is the appropriate model.

We also run a zero-inflated version of these models because we have an excess number of zeros in both variables (number of times looked and number of times licensed). In our sample of 285 patents, 85 patents have never been looked at and 132 have never been licensed (includes not looked at and looked at but not licensed). This zero-inflated models have two parts –a Poisson model and a logit model. The dependent variable in the Poisson part of the model is the number of times a patent has been looked at or licensed, conditional on being looked at or licensed, respectively. The separate inflation model which is a simple logit explains the excess zeros. Even though the zero inflated negative binomial model would be more appropriate, it doesn't converge and we report results from the zero inflated Poisson. A Vuong test shows that the zero inflated model is more appropriate than the regular Poisson.

The results from the zero-inflated Poisson models are similar to the previous three models. Interestingly, once controlled for the number of patents cited, a higher patent originality predicts more agreements, conditional on there being at least one agreement on the patent. The lack of lead inventor experience seems to determine whether a patent is never looked at or licensed but not the number of times it has been looked at or licensed. Our patent scope results contradict previous literature and seem to influence our dependent variables negatively and in some models at a statistically significant level (Lerner, 1994; Decheneaux, 2009; Gambardella et al. 2007)

The last three models are Cox hazard models. Here the dependent variable is time to first agreement. Positive results indicate that as the independent variable increases, so does the hazard of an agreement. Note, however, that for ease of interpretation beta coefficients are reported rather than hazard ratios. Hazard ratios can be calculated by raising e to the power of the reported coefficient. We report a regular Cox hazard model in the first column, then stratified by lead inventor and then by a shared frailty (the equivalent of a random effects model) where each group is identified by a lead inventor and includes said lead inventor's patents. Our results are similar to those from the previous models but are not statistically significant in the same manner.

The models in Table 1.4b are the same as the ones in Table 1.4a, except that the dependent variables are related to licensing – i.e. licensed, times licensed, and time to license.

The direction and the significance of the results are practically the same as well, except for the scope variable which is no longer negative and statistically significant. Another difference is that the "patent cites per year" variable is now positive and statistically significant in two of our models.

1.4.2 Patent-Agreement Level Models

Our patent-agreement level analyses are our main results. They test our hypothesis that technology proximity between a focal patent and a potential licensee's patent portfolio is a determinant of whether a license will take place. All our models are logistic regressions (logits) with a dependent variable – "deal" – that is equal to one if a license was signed and zero if a deal was not done. Our full model is the last one in the respective table and it includes all our control variables. We start with a logit of our dependent variable with only the respective proximity/similarity measures. We then try a bare-bones proximity and licensee variable model. For our remaining models we start again with the fit measure and add patent citation based measures such as forward cites, backward cites and scope. We then add patent variables constructed based on cites and IPC codes - originality measure, generality measure. Technology type controls are added next – i.e. a device indicator variable and mutually exclusive dummy variables based on IPC classification by sections. Age variables are included next - technology age at time of agreement and cohort dummies for each 5 year period since 1980 based on the patent priority date. Finally, we add licensee variables – the number of granted patents that the licensee had filed before the time of the agreement and the square of the number of such patents.

Our main models in Table 1.5a test our hypotheses above that firms are more likely to license inventions that are similar but not too similar. We operationalize our technology proximity measures using the cosine variables described above. This model includes our entire sample with 588 patent–agreement level observations. We replace the cosine measures with 0s in cases in which they are not defined because the agreement firm has no patents. In table 1.5b we exclude those observations with undefined cosine measures and are left with 424 observations. As seen from the tables, the sign and statistical significance of our proximity coefficients is unchanged.

We see in these models that a higher technological proximity between the firm and the AMC patent, measured at the IPC code subclass level, is more likely to be associated with a license i.e. a "deal." Holding subclass level proximity constant, however, a higher technological proximity measured at the main group level of the IPC codes is less likely to result in a deal.

This result is hard to comprehend because of the complexity of the measures. A stylized example may be useful to let us abstract from the details of the patent classes by using more easily understandable consumer product categories that may be protected by these patents. Note that this is a simplification for clarification purposes– our IPC subclass level and IPC main group levels do not necessarily correspond to the product categories described here, including their breath and generality.

Imagine a medical device company A that has 10 patents - 5 for metal and 5 for ceramic prosthetic hip implants. These technologies are all in the 'prosthetic hips' area at the broad level of measurement but different at the more granular level – 'type of prosthetic hip.' Another firm B also has 10 patents but all 10 for plastic hip implants. A third firm C, also a medical device company has patents for stents and catheters but none in the hip implants area. All three firms have signed confidentiality agreements showing interest in an AMC patent of a *plastic* hip implant. If we only looked at the broad level measures, both firm A and firm B would be more likely to license the new plastic hip implant patent that is offered by our AMC than firm C – their technology proximity measure to the patent offered is higher than that of firm C. Also, firm A would be as equally likely to license it as firm B as their proximity measure values are the same at this level.

Now let's look at firm A and firm B, at the more granular level of proximity, controlling for the higher level proximity. Based on our results we would expect that firm A is more likely to license the AMC patent than firm B. While firm B already has plastic hip implant patents, firm A has none, so at the granular level of measurement, firm B's portfolio is closer to the focal patent than firm A's portfolio which makes it less likely to license it.

These results are in fact intuitive. A match at the broad level (i.e. hip implants) is useful since it indicates that both firms may have specific complementary technologies to the focal patent. One example would be that both firms have their own technology that deals with specific implant shape (i.e. implant is hollow inside to enhance movement) or components needed to attach the implant. These broad level technologies would be likely relevant to implants of any materials. In general, their existing broad level complementary technologies will increase their marginal benefit of licensing any patent for which they are relevant. However, firm B already has its own *plastic* implant technology. If it licensed our focal patent, it would most likely be duplicating its own technology at a cost. Furthermore, choosing outside *plastic* hip technology over internally developed one may be harder given the lack of knowledge about the outside technology and the difficulty in transferring tacit knowledge (Polanyi, 1966; von Hippel, 1988; Agarwal, 2006). This implies that a higher similarity measured at the more granular, detailed level is likely to not result in a license.

To understand whether these results are different for different types of firms, we split our sample by firm R&D age and by firm size. Our goal is to separate firms that are based on different types of technology platforms - young biotech firms that do mostly large molecule drug development and older pharmaceutical firms specializing in small molecule drug discovery. Our samples are not entirely "clean" in the sense that there are also medical device and software companies in both groups. Furthermore, the distinction between biotech and pharmaceutical firms has grown blurrier over the years as pharmaceutical firms have also developed large molecule drug technologies and young biotechnology companies who started based on one specific invention have developed related expertise. Interaction effects would have been better for these exercises but they would have complicated interpretation significantly because we have two different variables we want to interact with multiple size and age variables.

In Tables 1.6a through 1.6c we split our sample in different groups by age - firms with an R&D age of less than 10 years, those with an R&D age between 10 and 20 years and firms older than 20 years. We note that our results have the same sign as in the full sample models but are not statistically significant in many of the models in the first age group, especially the measures based on more granular IPC code slices at the IPC main group level. Our results are, however very strong in terms of both size and statistical significance in the 10-20 year old firms group. They are also strong in a few of the models in the older firm group. Note however, that the addition of technology age and technology cohort variables changes the value and significance of the proximity measures indicating that they are related. It is also interesting to point out here that technology age is negative and significant among medium aged firms (10-20 years of R&D) and positive and significant for older firms (20+ years of R&D). This implies that older firms are more likely to license older and more established technologies, while younger firms may be more willing to undertake risks. This result is not driven by inventor startups, however, as those would have no patents at the time of licensing and would not be in any of the age groups. Note that several of our control variables for IPC code section and cohort drop resulting in smaller sample sizes for some of the models.

We then go on to split our samples by size. By looking at firm names and the number of patents that they own at time of licensing we find out that most biotechnology firms have fewer than 500 patents and most big pharmaceutical companies have more than 500 patents. Again, this division is not "clean" in the sense discussed above. The sign of our major results hold again. Interestingly, the results are only statistically significant for the cosine measure at the subclass level in the smaller sized group and only for the cosine measure at the more granular, main group level for the larger firms. This implies that finding a closely related technology is more important for smaller firms while not duplicating efforts may be more important for larger firms. Note, however, that controlling for size in the sample of firms with more than 500 patents gives us very large and highly significant coefficients for both proximity measures in the full model.

The next set of models in Table 1.8 includes only the first agreement that is signed for a patent whether it is a "deal" or "no deal." We are interested in these results because we are concerned that whether the first agreement is a "deal" or "no deal" may signal patent quality and may influence future licenses, especially non-exclusive licenses of which there may be potentially many per patent. An exclusively licensed invention, on the other hand, takes the patent off the market. Of the 200 patents that have at least one agreement, 6 are excluded because of licensee issues (discussed in the data section) and we are left with 194 patent-first agreement pairs. Our technology proximity results from the previous models still hold in this sample and are statistically significant indicating that the results are robust and are not driven by a few patents that have been licensed multiple times since each patent appears only once in this dataset.

1.6 Conclusion

In this paper we addressed a gap in the literature on markets for technology by taking a close look at the demand for technology. While this has been attempted in previous papers, our unique dataset that includes not only firms that licensed technologies but also showed interest in them but did not license provides an important control group for our description of the structure of such markets. We showed that proximity matters in the technologies firms decided to license. Our identification comes from variation within a group that showed at least a threshold level of interest in the technology by contacting the licensing office and signing a confidentiality agreement. Future research will expand on this by identifying a larger population of potential buyers in this market based on some other measure of interest – we currently do not include informal channels through which information may have been obtained or inquiries that did not result in signing of a confidentiality agreement.

We also contribute to the literature on measurement of technology proximity by using a new patent statistic – the international patent class which with its nested structure allows for proximity measurement at the broad as well as granular level between different (portfolios of) patents. We also improved on existing measures by including multiple classes rather than just one, resulting in more robust results (Benner and Weldfogel, 2008). Further comparison and validation of these new measures is in order.

Ultimately the real question is whether these technologies make it to the product market once they are licensed and how the technology proximity, either at the broad or the granular level influences that outcome. It would be interesting to know whether in-licensed technologies that are very close to the licensee's in-house developed technology are strategically shelved or perhaps not absorbed by the firm due to behavioral resistance to outside innovations, the so called "not-invented-here" syndrome (Katz and Allen, 1982; Thursby and Thursby, 2004). We view this paper as a first step in this exciting direction.

1.7 Bibliography

Agarwal, AK (2006), "Engaging the Inventor: Exploring Licensing Strategies for University Inventions and the Role of Latent Knowledge" *Strategic Management Journal*, 27 (1), pp. 63-79

Anand B.N. and T. Khanna (2000),"The structure of licensing contracts", Journal *of Industrial Economics*, 48 (1), pp. 103 – 135

Anton, J.J. and D.A. Yao (1994), "Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights." American Economic Review, vol. 84 pp.190-209.

Arora, A., Fosfuri, A., and Gambardella, A. (2001) Markets for Technology – The Economics of Innovation and Corporate Strategy, Cambridge, MA: MIT Press

Arora, A. and A. Gambardella (1994), 'Evaluating technological information and utilizing it: scientific knowledge, technological capability, and external linkages in biotechnology,' *Journal of Economic Behavior & Organization*, 24(1), pp. 91–114

Arora, A. and A. Gambardella (2010), "Ideas for Rent: An Overview of Markets for Technology," *Industrial and Corporate Change*, vol. 19 (3) pp. 775-803

Arora, A. and A. Fosfuri (2003), 'Licensing the market for technology,' *Journal of Economic Behavior and Organization*, 52(2), pp. 277–295.

Arrow, K. (1962), "Economic Welfare and the Allocation of Resources for Invention," in: *The Rate and Direction of Inventive Activity: Economic and Social Factors*, National Bureau of Economic Research, Inc., pp.609-626

Athreye, S. and J. Cantwell (2007), 'Creating competition? Globalization and the emergence of new technology producers,' *Research Policy,* 36(2), pp. 209–226

Atuahenegima, K (1993), "Buying Technology for Product Development in Smaller Firms", *Industrial Marketing Management*, 22 (3), pp. 223-232

Atuahenegima, K. (1992), Inward Technology Licensing As an Alternative to Internal Research-and-Development in New Product Development - a Conceptual-Framework, *Journal of Product Innovation Management*, 9(2), pp.156-167

Atuahenegima, K. and P. Patterson (1993), "Managerial Perceptions of Technology Licensing as an Alternative to Internal Research-and-Development in New Product Development - an Empirical-Investigation," *R & D Management*, 23 (4), pp. 327-336

Benkard, L. (2000), "Learning and Forgetting: The Dynamics of Aircraft Production", *American Economic Review,* 90 (4), pp. 1034 – 1054

Benner, M. and J. Waldfogel (2008), "Close to you? Bias and precision in patent-based measures of technological proximity", *Research Policy*, 37(9), pp. 1556-1567

Bessy, C and E. Brousseau, (1998) "Technology licensing contracts features and diversity", *International Review of Law and Economics*, 18(4) pp. 451-489

Breschi S., F. Lissoni F and F. Malerba (2003), "Knowledge-relatedness in firm technological diversification," *Research Policy,* 32 (1), pp. 69-87

Bresnahan, T. and M. Trajtenberg (1995), "General purpose technologies 'Engines of growth'?" *Journal of Econometrics*, 65(1), pp. 83-108

Cassiman B. and R. Vergeulers (2006), In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition, *Management Science*, 52 (1), pp. 68-82

Caves, R., H. Crookell and JP Killing (1983), "The Imperfect Market for Technology Licenses," *Oxford Bulletin of Economics and Statistics*, 45 (3), pp.246-267

Chatterji, A.K., K. Fabrizio, W. Mitchell and K.A. Schulman (2008), Physician-Industry Cooperation in the Medical Device Industry, *Health Affairs*, 27(6), pp. 1532-1543

Cockburn, I.M., M.J. MacGarvie and E. Muller (2010) Patent thickets, licensing and innovative performance, *Industrial And Corporate Change*, 19 (3), pp. 899-925

Cohen, W. and D. Levinthal (1990), "Absorptive Capacity: A New Perspective on Learning and Innovation". *Administrative Science Quarterly,* 35(1), pp. 128-152

Dechenaux E., B. Goldfarb, S. Shane and M. Thursby (2008), Appropriability and Commercialization: Evidence from MIT Inventions, *Management Science*, 54 (5), pp. 893-906

Ford, D. (1988), "Develop Your Technology Strategy," Long Range Planning, 11(5), pp. 85 -95

Gale, D. and L. S. Shapley (1962), "College Admissions and the Stability of Marriage", *American Mathematical Monthly*, 69 (1), pp.9-14

Gambardella, A., P. Giuri and A. Luzzi (2007), “The market for patents in Europe,” Research Policy, 36(8), pp.1163–1183

Gans, J., D. H. Hsu and S. Stern (2008), “The impact of uncertain intellectual property rights on the market for ideas: evidence from patent grant delays,” Management Science, 54(5), pp.982–997.

Gans, J. and S. Stern (2010), “Is There a Market for Ideas”, *Industrial and Corporate Change*, 19(3), pp. 805-837

Haagedorn, J., Lorenz-Orlean S. and H. Kranenburg (2009), “Technology Transfer: Partnership Embedded Licensing Agreements”, Industrial and Corporate Change, 18 (3), pp. 529-550

Hall, B., A. Jaffe and M. Trajtenberg (2001), “The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools,” NBER Working Paper No. 8498

Heller, MA and RS Eisenberg (1998), Can Patents Deter Innovation? The Anticommons in Biomedical Research, *Science, v*ol. 280 (5364), pp. 698-701

Henderson, R., A. Jaffe & M. Trajtenberg (1998), *"*Universities as a Source of Commercial Technology*:* A Detailed Analysis of University Patenting, 1965-1988", *Review of Economics and Statistics,* Vol. 80(1), pp. 119-127

Higgins, M. J. and D. Rodriguez (2006), “The outsourcing of R&D through acquisitions in the pharmaceutical industry,” *Journal of Financial Economics,* 80(2), pp. 351–383

Jaffe, Adam B. (1986) “Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits and Market Value,” American Economic Review, 76, pp. 984–1001

Katz, R. and T. Allen (1982), Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D Project Groups, *R&D Management*, 12 (1), pp. 7-20

Killing, JP (1978), “Diversification through Licensing”, R&D Management, 8(3), pp. 159 – 163

Kurokawa S. (1997), “Make-or-buy decisions in R&D: Small technology based firms in the United States and Japan,” *IEEE Transactions on Engineering Management,* 44 (2), pp. 124-134

Laursen, K; M, Leone, and S. Torrisi, (2010), "Technological exploration through licensing: new insights from the licensee's point of view," *Industrial and Corporate Change*, 19 (3) pp. 871 - 897

Lerner, J (1994), The Importance Of Patent Scope - An Empirical-Analysis , *Rand Journal Of Economics*, 25 (2), pp. 319-333

Lichtenthaler, U. (2007), "The drivers of technology licensing: an industry comparison," California Management Review, 49(4), pp. 67–89.

Lichtenthaler, U. and H. Ernst (2007), "External technology commercialization in large firms: results of a quantitative benchmarking study," R&D Management, 37(5), pp. 383–397

Lowe J. and P. Taylor (1998) R & D and Technology Purchase through License Agreements: Complementary Strategies and Complementary Assets, *R & D Management*, 28 (4), pp. 263-278

Mehta,A., M. Rysman and T. Simcoe (2010),"Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion", *Journal of Applied Econometrics,* 25 (7), pp. 1179-1204

Mowery D., J. Oxley and B. Silverman (1996), "Strategic Alliances And Interfirm Knowledge Transfer," *Strategic Management Journal*, 17 (SI), pp.77-91

Narin F., E. Noma and R. Perry (1987) "Patents As Indicators of Corporate Technological Strength ," *Research Policy* , 16 (2-4), pp. 143-155

Nerkar A. and S. Shane (2007), "Determinants of invention commercialization: An empirical examination of academically sourced inventions," *Strategic Management Journal,* 28 (11), pp.1155-1166

Patel, P and K. Pavitt (1994), "The Continuing, Widespread (and Neglected) Importance of Improvements In Mechanical Technologies ," *Research Policy,* 23, pp. 533-545

Pisano, G. (1990) "The R&D Boundaries of the Firm: An Empirical Analysis," *Administrative Science Quarterly,* 35 (1) pp. 153-176

Pitkethly, R.H. (2001), "Intellectual property strategy in Japanese and UK companies: patent licensing decisions and learning opportunities", *Research Policy*, 30(3), pp. 425–442

Polanyi, M (1966), The Tacit Dimension, Garden City, N.Y. : Doubleday

Razgaitis, R. (2004), “US/Canadian licensing in 2003: survey results,” *Journal of the Licensing Executive Society*, 34(4), pp.139–151

Robbins, C. (2006), ‘Measuring payments for the supply and use of intellectual property,’ Bureau of Economic Analysis, U.S. Department of Commerce: Washington, DC

Rosenberg, N. (1990). “Why Do Firms Do Basic Research (With Their Own Money)?” *Research Policy, 19*(2), pp. 165-174.

Roth, A. E. (2008) "What have we learned from market design?" Hahn Lecture*, Economic Journal*, 118 (March), pp. 285–310

Sheehan, J., C. Martinez and D. Guellec (2004), ‘Understanding business patenting and licensing: results of a survey,’ OECD: Paris.

Silverman B. (1999), “Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics,” *Management Science*, 46 (8), pp. 1109-1124

Teece, D. (1986), “Profiting from Technological Innovation: Implications for integration, collaboration, licensing and public policy”, *Research Policy*, 15 (6), pp. 285-305

Thursby J.G. and M.C. Thursby (2004), “Are faculty critical? Their role in university-industry licensing”, *Contemporary Economic Policy*, 22 (2), pp. 162-178

Trajtenberg, M. (1990) A Penny for your Quotes - Patent Citations and the Value of Innovations, *RAND Journal of Economics*, 21 (1), pp. 172-187

Tsai, K. and J.Wang, (2007), Inward technology licensing and firm performance: a longitudinal study, *R & D Management*, 37 (2), pp. 151-160

Tsai, K. and J.Wang, (2009) “External technology sourcing and innovation performance in LMT sectors: An analysis based on the Taiwanese Technological Innovation Survey”, *Research Policy*, 38 (3), pp.518–526

von Hippel, E. (1976) "The Dominant Role of Users in the Scientific Instrument Innovation Process," *Research Policy,* 5(3), pp. 212-39

von Hippel, E. (1988), The Sources of Innovation , New York: Oxford University Press

Zeckhauser R. (1996), “The Challenge of Contracting for Technological Information," *Proceedings of the National Academy of Sciences*, 93, pp.12743-12748

Zuniga, M. P. and D. Guellec (2008), ‘Survey on patent licensing: initial results from Europe and Japan,’ OECD: Paris.

Table 1.1a This table contains descriptive statistics for the AMC patents that are to be licensed. Two sided t-tests of difference in means between "Never Looked At" and "Licensed at Least Once"; significance indicated in the "Never Looked At" mean column; Two sided t-test of difference in means between "Looked At, Not Licensed" and "Licensed At Least Once"; significance indicated in the "Looked At, Not Licensed" mean column; *** p<0.01, ** p<0.05, * p<0.1

	Never Looked At					Looked At, Not Licensed					Licensed At Least Once				
	N	Mean	Std Dev	Min	Max	N	Mean	Std Dev	Min	Max	N	Mean	Std Dev	Min	Max
Lead Inventor Experience	**85**	4.61***	3.43	1	15	**48**	6.02	4.61	1	17	**152**	7.19	4.77	1	25
Patent Scope	**85**	2.43***	1.38	1	9	**48**	2.87	1.48	1	6	**152**	3.71	3.49	1	19
Cites Per Year	**85**	0.78	1.18	0	6	**48**	0.33**	0.59	0	3	**152**	0.96	1.83	0	19
Cites First Two Years	**85**	0.42	0.90	0	6	**48**	0.33	0.88	0	4	**152**	0.47	0.85	0	5
Originality	**85**	0.64	0.32	0	1	**48**	0.67	0.29	0	1	**152**	0.68	0.23	0	1
Generality	**62**	0.62	0.20	0	1	**25**	0.72***	0.22	0	1	**126**	0.57	0.24	0	1
Number of Cited Patents	**85**	6.57***	6.97	0	29	**48**	6.83**	7.27	0	37	**152**	10.16	10.19	0	54
Share of Agreements that are Licenses											**152**	0.79	0.28	0.125	1
Times Looked						**48**	2.22**	1.85	1	8	**152**	3.24	3.12	1	15
Times Licensed											**152**	2.02	1.66	1	11
Time to First Agreement						**48**	2.75	2.45	0.00	7.98	**152**	3.44	3.68	0.00	12.76
Time to First License											**152**	4.20	3.84	0.00	18.48
Age of Oldest Cited Patent	**66**	15.69**	11.48	2.34	55.56	**45**	17.29	12.66	1.36	57.55	**140**	21.55	17.34	1.09	74.81
Mean Age of Cited Patents	**66**	7.53***	3.95	1.94	18.73	**45**	8.71	4.31	1.36	16.99	**140**	9.94	4.84	1.09	25.21
Median Age of Cited Patents	**66**	6.38***	3.52	1.74	14.46	**45**	7.74	3.92	1.36	16.99	**140**	8.68	4.25	1.09	22.84

Table 1.1b This table contains descriptive statistics for the AMC patents that are to be licensed. Two sided t-tests of difference in means between "Never Looked At" and "Licensed at Least Once"; significance indicated in the "Never Looked At" column; Two sided t-test of difference in means between "Looked At, Not Licensed" and "Licensed At Least Once"; significance indicated in the "Looked At, Not Licensed" column; *** p<0.01, ** p<0.05, * p<0.1

Group	Never Looked at (N=85)	Looked At, Not Licensed (N=48)	Licensed (N=152)
	Proportion of "Never Looked At" Patents in Group	Proportion "Looked At, Not Licensed" Patents in Group	Proportion of "Licensed" Patents in Group
IPCs in Section A Only	0.36	0.35	0.38
IPCs in Section C Only	0.14	0.06	0.14
IPCs in Section G Only	0.22***	0.14**	0.05
IPCs in Sections A and G Only	0.02	0.04	0.05
IPCs in Sections A and C Only	0.18*	0.19	0.30
IPCs in Sections A, C and G Only	0.01*	0.06	0.07
Patent Priority Date in 1981-85	0.07	0.00*	0.07
Patent Priority Date in 1986-90	0.28	0.02***	0.25
Patent Priority Date in 1991-95	0.23***	0.08***	0.43
Patent Priority Date in 1996-00	0.30*	0.66***	0.20
Patent Priority Date in 2001-04	0.10*	0.22***	0.05
Device	0.38***	0.29	0.20

Table 1.2a This table contains descriptive statistics for the agreement-patent level data. Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. Two sided t-tests of difference in means between "Deal " and "No Deal" ; significance indicated in the "No Deal" mean column; *** p<0.01, ** p<0.05, * p<0.1

Variable	No Deal					Deal				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Cosine Subclass Level†	287	0.24**	0.31	0	1	301	0.30	0.31	0	1
Cosine Subclass Level (no replacement)††	198	0.35	0.32	0	1	226	0.39	0.30	0	1
Within Section Cosine, Group Level†	287	0.22	0.31	0	1	301	0.22	0.28	0	1
Within Section Cosine, Group Level (no replacement)††	198	0.33	0.32	0	1	226	0.29	0.28	0	1
Patent Scope	287	3.06*	1.75	1	10	301	3.45	2.94	1	19
Patent Citations Received Per Year	287	0.57***	0.74	0	3.67	301	0.95	1.79	0	18.91
Patent Citations First Two Years	287	0.38	0.74	0	5	301	0.45	0.87	0	5
Number of Patents Cited	287	10.88	10.66	0	54	301	10.53	10.22	0	54
Patent Originality	287	0.71	0.22	0	1	301	0.71	0.20	0	1
Patent Generality	194	0.60	0.26	0	1	246	0.58	0.24	0	1
Device	287	0.15	0.36	0	1	301	0.12	0.33	0	1
Technology Age in Years	287	5.02***	4.38	0	18.11	301	6.64	4.75	0	19.35
Lead Inventor Experience	287	7.85	5.10	1	22	301	7.27	4.52	1	25
Firm R&D age	198	17.04***	11.47	0.06	43.27	226	21.20	12.01	0.08	46.17
Number of Firm Patents	287	985.01***	2657.38	0	16392	301	2204.26	4227.66	0	15266
Numbers of AMC Patents Under Agreement	287	2.75***	1.70	1	6	301	4.20	4.36	1	19

†Observations for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0.

††Observations for which a cosine measure was not defined were excluded from this calculation.

Table 1.2b This table contains descriptive statistics for the agreement-patent level data. Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. Two sided t-tests of difference in means between "Deal " and "No Deal" ; significance indicated in the "No Deal" mean column; *** p<0.01, ** p<0.05, * p<0.1

	No Deal (N=287)	Deal (N=301)
Group	Proportion of "No Deal" Patent Observations in Group	Proportion of "Deal" Patent Observations in Group
Firm Has No Patents/Firm R&D age undefined	0.31*	0.25
Firm Has >0 and < 500 Patents	0.47	0.43
Firm Has >500 Patents	0.22***	0.32
Firm R&D age >0 and < 10 yrs Old	0.25***	0.13
Firm R&D age >10 and <20 yrs	0.18**	0.27
Firm R&D age >20 yrs	0.26**	0.35
Patent Priority Date in 1981-85	0.13	0.13
Patent Priority Date in 1986-90	0.13***	0.25
Patent Priority Date in 1991-95	0.09***	0.36
Patent Priority Date in 1996-00	0.53***	0.25
Patent Priority Date in 2001-04	0.13***	0.02
IPCs in Section A Only	0.42	0.40
IPCs in Section C Only	0.07***	0.15
IPCs in Section G Only	0.05	0.03
IPCs in Sections A and C Only	0.25	0.31
IPCs in Sections A and G Only	0.04	0.06
IPCs in Sections A, C and G Only	0.10**	0.05

Table 1.3 This table contains descriptive statistics for the agreement level data. Each observation corresponds to a unique agreement. Two sided t-tests of difference in means between "Deal " and "No Deal" ; significance indicated in the "No Deal" mean column; *** p<0.01, ** p<0.05, * p<0.1

	No Deal					Deal				
Variable	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Number of Firm Patents	153	1028.43	2740.24	0	16392	142	1328.67	3039.97	0	15266
Firm Technoology Age At Time of Agreement	153	12.44	12.40	0	43	142	13.33	12.79	0	46
Number of Patents on Agreements	153	1.88	1.28	1	6	142	2.12	2.11	1	19
Agreement Level Cosine, Main Group Level†	153	0.20	0.27	0	0.96	142	0.18	0.26	0	0.94
Agreement Level Cosine, Main Group Level (no replacement)††	114	0.26	0.28	0	0.96	99	0.26	0.27	0	0.94
Agreement Level Cosine, Subclass Level†	153	0.27	0.32	0	1	142	0.29	0.33	0	1
Agreement Level Cosine, Subclass Level (no replacement)††	114	0.36	0.33	0	1	99	0.42	0.32	0	1

†Observations for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0.
††Observations for which a cosine measure was not defined were excluded from this calculation.

Table 1.4a This table contains models of patent consideration for licensing i.e. "looks at patent" based on patent characteristics. P-values reported under coefficients.

	LOGIT	**Poisson**	**Negative Binomial Regression**	**Zero Inflated Poisson**		**Cox Hazard Model**	**Cox Hazard Model** Stratified by Lead Inventor	**Cox Hazard Model** Frailty by Lead Inventor
VARIABLES	Looked (looked=1)	Times looked	Times looked	Times looked	Inflate (looked=0)	Time to Look	Time to Look	Time to Look
Number of Patents Cited	0.0837***	0.0311***	0.0346***	0.0225***	-0.110*	0.013	0.020	0.0219*
	0.001	0.000	0.000	0.000	0.096	0.113	0.276	0.083
Patent Citations Received Per Year	0.023	0.042	0.058	0.039	-0.737	0.006	0.342*	-0.009
	0.818	0.134	0.230	0.162	0.292	0.875	0.0621	0.903
Patent Scope	0.103	-0.0388**	-0.037	-0.0456**	-0.16	-0.0952***	-0.007	-0.031
	0.263	0.045	0.234	0.019	0.405	0.005	0.919	0.515
Lead Inventor Experience	0.118***	0.0349***	0.0380**	0.011	-0.293***	0.000	0.115	0.030
	0.002	0.000	0.011	0.265	0.00264	0.978	0.144	0.348
Patent Originality	0.318	0.208	0.309	0.560***	2.399	0.0381	-0.976	-0.107
	0.582	0.261	0.298	0.009	0.186	0.909	0.189	0.833
Device	-1.047**	-0.824***	-0.844***	-0.972***	-12.59	-0.125	-34.3	-0.523
	0.018	0.000	0.000	0.000	0.977	0.58	1.000	0.215
IPCs in Section A Only	0.425	0.234	0.289	0.409*	1.166	0.370	-0.099	-0.273
	0.558	0.305	0.399	0.093	0.423	0.325	0.898	0.554
IPCs in Section C Only	0.398	-0.048	0.024	0.161	0.208	-0.040	-0.096	-1.226**
	0.615	0.851	0.949	0.553	0.891	0.925	0.916	0.031
IPCs in Section G Only	-0.250	-0.620**	-0.457	-0.565*	-12.69	-0.287	-0.550	-0.826
	0.761	0.035	0.271	0.059	0.988	0.521	0.595	0.164
IPCs in Sections A and C Only	0.404	0.152	0.166	0.183	-0.617	-0.060	-0.151	-1.309***
	0.576	0.507	0.626	0.449	0.678	0.875	0.841	0.006
IPCs in Sections A and G Only	1.318	0.795***	0.826*	0.680**	-24.07	0.397	-0.445	-0.565
	0.207	0.005	0.061	0.019	1	0.424	0.729	0.404
IPCs in Sections A, C and G Only	1.827	0.843***	0.946**	0.779***	-0.836	0.117	-0.566	-0.6
	0.147	0.002	0.031	0.006	0.633	0.817	0.643	0.407
Patent Priority Date in 1981-85	-0.901	0.646***	0.305	0.749***	3.522	0.346	-1.815	0.614
	0.245	0.001	0.394	0.001	0.111	0.414	0.285	0.378
Patent Priority Date in 1986-90	-0.481	0.062	-0.132	-0.006	1.798	-0.286	-3.095**	-1.127**
	0.389	0.743	0.633	0.976	0.376	0.361	0.041	0.029
Patent Priority Date in 1991-95	0.163	-0.085	-0.139	-0.165	1.193	-0.711**	-2.789**	-1.177**
	0.765	0.642	0.601	0.390	0.565	0.023	0.050	0.021
Patent Priority Date in 1996-00	-0.001	0.413**	0.415	0.461**	2.454	0.307	0.105	0.177
	0.998	0.016	0.109	0.010	0.216	0.298	0.924	0.708
Constant	-0.923	-0.025	-0.132	0.109	-2.673			
	0.380	0.939	0.785	0.757	0.323			
lnalpha			-0.457***					
			0.006					
Observations	285	285	285	285	285	200	200	200
Groups								86

*** p<0.01, ** p<0.05, * p<0.1

Table 1.4b This table contains models of licensing based on patent level dependent characteristics. P-values under coefficients.

	LOGIT	**Poisson**	**Negative Binomial Regression**	**Zero Inflated Poisson**		**Cox Hazard Model**	**Cox Hazard Model** Stratified by Lead Inventor	**Cox Hazard Model** Frailty by Lead Inventor
VARIABLES	Licensed licensed=1	Times licensed	Times licensed	Times licensed	Inflate licensed=0	Time to License	Time to License	Time to License
Number of Patents Cited	0.0926***	0.0307***	0.0305***	0.008	-0.286***	0.010	0.036	0.006
	0.000	0.000	0.000	0.224	0.001	0.280	0.140	0.694
Patent Citations Received Per Year	0.048	0.0695**	0.0819*	0.038	-2.630**	-0.027	0.339	-0.060
	0.650	0.031	0.066	0.288	0.012	0.567	0.125	0.394
Patent Scope	0.143	0.005	0.007	-0.020	-0.397	-0.021	0.031	0.029
	0.111	0.827	0.823	0.435	0.164	0.567	0.685	0.566
Lead Inventor Experience	0.103***	0.0330**	0.0304*	-0.008	-0.193**	-0.0374*	0.269**	-0.045
	0.003	0.015	0.076	0.596	0.036	0.083	0.040	0.200
Patent Originality	0.384	0.389	0.417	0.267	-1.486	-0.101	0.299	-0.617
	0.527	0.150	0.215	0.441	0.297	0.830	0.712	0.266
Device	-1.014**	-1.013***	-0.973***	-1.181***	-0.444	-0.210	-39.300	-0.775*
	0.029	0.000	0.000	0.000	0.684	0.454	1.000	0.082
IPCs in Section A Only	2.641***	1.310**	1.331**	1.022*	-0.841	-2.254***	0.267	-1.915*
	0.004	0.011	0.016	0.067	0.695	0.005	0.830	0.063
IPCs in Section C Only	3.429***	1.375***	1.425**	0.710	-4.456*	-2.894***	0.676	-2.987***
	0.001	0.010	0.014	0.225	0.084	0.000	0.623	0.006
IPCs in Section G Only	1.869*	0.619	0.653	0.570	-0.271	-2.799***	0.602	-2.048*
	0.075	0.291	0.304	0.368	0.914	0.002	0.720	0.079
IPCs in Sections A and C Only	3.083***	1.248**	1.321**	0.566	-5.361**	-3.264***	-0.390	-3.839***
	0.001	0.015	0.017	0.312	0.024	0.000	0.745	0.000
IPCs in Sections A and G Only	3.711***	2.097***	2.202***	1.097*	-32.300	-2.662***	0.631	-3.023***
	0.001	0.000	0.000	0.067	1.000	0.002	0.712	0.008
IPCs in Sections A, C and G Only	3.355***	1.216**	1.234*	0.556	-4.815**	-3.808***	-0.044	-2.919**
	0.004	0.039	0.058	0.379	0.049	0.000	0.978	0.011
Patent Priority Date in 1981-85	1.438*	1.873***	1.787***	1.644***	1.010	-0.316	-19.160	-0.946
	0.080	0.000	0.000	0.001	0.689	0.542		0.253
Patent Priority Date in 1986-90	1.743***	1.471***	1.436***	1.191**	-0.864	-0.212	-19.67***	-1.096
	0.005	0.000	0.001	0.014	0.692	0.624	0.000	0.108
Patent Priority Date in 1991-95	2.239***	1.518***	1.559***	1.136**	-2.809	-0.831*	-18.09***	-0.898
	0.000	0.000	0.000	0.017	0.196	0.064	0.000	0.183
Patent Priority Date in 1996-00	0.401	1.110***	1.092**	1.371***	2.358	0.077	-20.93***	-0.617
	0.493	0.006	0.014	0.004	0.247	0.862	0.000	0.365
Constant	-5.715***	-3.190***	-3.251***	-1.472*	7.626**			
	0.000	0.000	0.000	0.073	0.033			
lnalpha					-0.893***			
					0.002			
Observations	285	285	285	285	285	152	152	152
Groups								62

*** p<0.01, ** p<0.05, * p<0.1

Table 1.5a This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. **Firms for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0 in these models**. P-values under coefficients. Robust Standard Errors

	5.1a	5.2a	5.3a	5.4a	5.5a	5.6a	5.7a	5.8a
Cosine Subclass Level	1.157***	1.285***	1.127***	1.977***	1.361***	1.238***	1.085**	1.276***
	0.006	0.002	0.008	0.001	0.003	0.008	0.022	0.008
Within Section Cosine, Group Level	-0.921**	-1.056**	-0.968**	-2.079***	-1.259***	-1.031**	-0.856*	-1.042**
	0.032	0.014	0.028	0.001	0.010	0.037	0.094	0.041
Patent Scope			0.050	0.051	0.112***	0.105***	0.271***	0.260***
			0.122	0.229	0.005	0.010	0.000	0.000
Number of Patents Cited			0.001	0.009	0.006	0.004	-0.004	-0.006
			0.948	0.375	0.526	0.637	0.763	0.643
Patent Citations Received Per Year			0.373***	0.218**	0.404***	0.378***	0.264**	0.272**
			0.001	0.043	0.000	0.001	0.037	0.033
Patent Originality				-0.195	1.014**	0.853	1.310**	1.291**
				0.717	0.047	0.101	0.028	0.032
Device					-0.332	-0.308	-0.180	-0.110
					0.274	0.321	0.612	0.765
IPCs in Section A Only					1.539***	1.492**	1.746***	1.570**
					0.009	0.012	0.009	0.018
IPCs in Section C Only					2.551***	2.454***	3.264***	3.199***
					0.000	0.000	0.000	0.000
IPCs in Section G Only					0.768	0.828	1.588*	1.519*
					0.344	0.286	0.055	0.076
IPCs in Sections A and C Only					1.573***	1.408**	2.133***	1.923***
					0.008	0.017	0.001	0.004
IPCs in Sections A and G Only					1.948***	1.737**	2.096**	1.889**
					0.005	0.013	0.013	0.020
IPCs in Sections A, C and G Only					0.266	-0.134	-0.552	-0.605
					0.701	0.852	0.510	0.464
Technology Age in Years						0.0760***	-0.005	-0.006
						0.001	0.837	0.812
Patent Priority Date in 1981-85							1.888***	2.038***
							0.004	0.002
Patent Priority Date in 1986-90							2.088***	2.122***
							0.000	0.000
Patent Priority Date in 1991-95							2.762***	2.793***
							0.000	0.000
Patent Priority Date in 1996-00							-0.075	-0.063
							0.884	0.906
Lead Inventor Experience	-0.022	-0.015	-0.019	-0.033	-0.017	-0.019	-0.037	-0.027
	0.204	0.413	0.297	0.132	0.398	0.340	0.120	0.263
Number of Firm Patents		0.000156*						0.000265**
		0.065						0.013
Number of Firm Patents Squared		0.000						-1.78e-08**
		0.557						0.030
Patent Generality				-0.354				
				0.428				
Constant	0.110	-0.131	-0.325	0.347	-2.805***	-2.965***	-4.666***	-4.709***
	0.526	0.484	0.170	0.586	0.000	0.000	0.000	0.000
Observations	588	588	588	440	588	588	588	588

*** p<0.01, ** p<0.05, * p<0.1

Table 1.5b This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. **Firms for which a cosine measure was not defined because the firm has no patents are excluded from these models.** P-values under coefficients. Robust Standard Errors

	5.1b	5.2b	5.3b	5.4b	5.5b	5.6b	5.7b	5.8b
Cosine Subclass Level	1.039**	1.430***	0.928**	1.905***	1.434***	1.149**	1.234**	1.683***
	0.019	0.001	0.040	0.003	0.009	0.040	0.034	0.006
Within Section Cosine, Group Level	-0.983**	-1.033**	-1.046**	-2.287***	-1.508***	-1.193**	-1.211**	-1.375**
	0.023	0.017	0.019	0.000	0.006	0.034	0.042	0.022
Patent Scope			0.0694*	0.054	0.108**	0.0879*	0.324***	0.308***
			0.057	0.293	0.019	0.063	0.000	0.000
Number of Patents Cited			-0.003	0.009	0.012	0.008	-0.008	-0.010
			0.778	0.436	0.282	0.486	0.599	0.525
Patent Citations Received Per Year			0.440***	0.387**	0.541***	0.492***	0.356**	0.383**
			0.001	0.016	0.000	0.000	0.025	0.019
Patent Originality				0.729	1.492**	1.199*	1.755**	1.902**
				0.327	0.018	0.064	0.024	0.019
Device					-0.631	-0.605	-0.343	-0.235
					0.100	0.121	0.408	0.583
IPCs in Section A Only					1.773**	1.682**	2.107**	1.953**
					0.037	0.045	0.024	0.042
IPCs in Section C Only					2.723***	2.596***	3.492***	3.495***
					0.003	0.004	0.000	0.001
IPCs in Section G Only					0.364	0.555	1.529	1.358
					0.783	0.645	0.198	0.288
IPCs in Sections A and C Only					1.854**	1.673**	2.591***	2.374**
					0.031	0.048	0.007	0.016
IPCs in Sections A and G Only					2.922***	2.746***	3.320***	3.135***
					0.004	0.006	0.006	0.009
IPCs in Sections A, C and G Only					0.446	0.102	-0.730	-0.746
					0.653	0.922	0.564	0.562
Technology Age in Years						0.104***	-0.010	-0.015
						0.000	0.778	0.687
Patent Priority Date in 1981-85							1.721**	1.847**
							0.018	0.014
Patent Priority Date in 1986-90							1.819***	1.708**
							0.006	0.011
Patent Priority Date in 1991-95							2.170***	2.148***
							0.000	0.000
Patent Priority Date in 1996-00							-0.747	-0.792
							0.190	0.179
Lead Inventor Experience	-0.016	-0.008	-0.006	-0.009	-0.008	-0.005	-0.016	-0.005
	0.432	0.727	0.785	0.754	0.748	0.855	0.607	0.881
Number of Firm Patents		0.000184**						0.000268**
		0.039						0.019
Number of Firm Patents Squared		0.000						-1.70e-08**
		0.418						0.047
Patent Generality				-0.874*				
				0.0913				
Constant	0.168	-0.304	-0.337	-0.233	-3.500***	-3.598***	-5.053***	-5.342***
	0.435	0.226	0.250	0.766	0.001	0.001	0.000	0.000
Observations	424	424	424	304	424	424	424	424

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6a Sample of firms that are older than 0 and younger than 10 years at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm technology age (i.e. time between the first patent that the firm filed and the agreement that it signed for the AMC patent) is smaller than 10 years. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	6.1a	6.2a	6.3a	6.4a	6.5a	6.6a	6.7a	6.8a
Cosine Subclass Level	0.763	0.818	0.959	1.529	2.012*	2.007*	2.919*	2.934*
	0.333	0.315	0.232	0.239	0.096	0.098	0.051	0.079
Within Section Cosine, Group	0.455	0.533	-0.158	-0.776	-1.631	-1.614	-3.708*	-3.305
	0.584	0.528	0.861	0.552	0.242	0.259	0.075	0.108
Patent Scope			-0.079	-0.159	-0.197	-0.198	-0.190	-0.200
			0.603	0.472	0.453	0.456	0.502	0.536
Number of Patents Cited			-0.029	-0.018	-0.036	-0.035	-0.048	-0.046
			0.236	0.517	0.311	0.343	0.341	0.368
Patent Cites Received Per Year			0.686**	0.525	1.143***	1.132***	1.211**	1.469**
			0.026	0.114	0.002	0.003	0.027	0.022
Patent Originality				0.459	0.634	0.615	-0.859	-0.627
				0.723	0.647	0.662	0.598	0.736
Device					1.327	1.311	1.660	1.009
					0.131	0.140	0.225	0.474
IPCs in Section A Only					15.49***	16.36***	18.64***	19.40***
					0.000	0.000	0.000	0.000
IPCs in Section C Only					15.67***	16.54***	20.15***	21.22***
					0.000	0.000	0.000	0.000
IPCs in Section G Only					13.95***	14.84***	17.67***	18.13***
					0.000	0.000	0.000	0.000
IPCs in Sections A and C Only					16.76***	17.63***	21.89***	22.24***
					0.000	0.000	0.000	0.000
IPCs in Sections A and G Only					15.17***	16.03***	0.006	1.687
					0.000	0.000	0.998	0.578
Lead Inventor Experience	-0.004	0.005	0.010	0.018	0.002	0.002	0.000	0.011
	0.919	0.904	0.845	0.773	0.976	0.980	0.998	0.863
Number of Firm Patents		0.062						0.107*
		0.103						0.091
Number of Firm Patents Squared		-0.001						-0.001
		0.299						0.183
Patent Generality				-1.616				
				0.236				
Technology Age in Years						0.006	-0.097	-0.085
						0.934	0.319	0.462
Patent Priority Date in 1986-90							18.24***	17.63***
							0.000	0.000
Patent Priority Date in 1991-95							1.112	0.963
							0.366	0.481
Patent Priority Date in 1996-00							-2.200**	-2.520**
							0.037	0.019
Constant	-1.077**	-1.661***	-1.114	-0.173	-17.47***	-18.34***	-19.05***	-20.76***
	0.042	0.007	0.137	0.906	0.000	0.000	0.000	0.000
Observations	110	110	110	76	105	105	98	98

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6b: Sample of firms that are older than 10 and younger than 20 years at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm technology age (i.e. time between the first patent that the firm filed and the agreement that it signed for the AMC patent) is larger than 10 and smaller than 20. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	6.1b	6.2b	6.3b	6.4b	6.5b	6.6b	6.7b	6.8b
Cosine Subclass Level	1.610**	1.787**	1.930**	2.634***	2.172**	3.022***	4.013***	4.397**
	0.021	0.046	0.013	0.002	0.010	0.002	0.003	0.027
Within Section Cosine, Group	-2.331***	-2.456**	-2.533***	-3.507***	-3.074***	-4.145***	-3.993***	-3.781***
	0.002	0.012	0.003	0.000	0.001	0.000	0.001	0.006
Patent Scope			-0.129	-0.158	0.023	0.087	0.542*	0.593
			0.333	0.364	0.901	0.624	0.092	0.118
Number of Patents Cited			0.039	0.053	0.031	0.024	0.0935*	0.111*
			0.159	0.138	0.425	0.585	0.056	0.075
Patent Cites Received Per Year			0.196	0.139	0.425	0.472*	0.404	0.482
			0.501	0.710	0.137	0.095	0.205	0.174
Patent Originality				-1.115	0.691	1.434	0.483	1.305
				0.502	0.733	0.505	0.815	0.571
Device					-0.541	-0.984	-2.839**	-2.752**
					0.459	0.210	0.017	0.046
IPCs in Section A Only					2.611**	3.501***	18.70***	17.98***
					0.032	0.005	0.000	0.000
IPCs in Section C Only					3.341***	3.901***	20.70***	20.42***
					0.010	0.003	0.000	0.000
IPCs in Sections A and C Only					1.288	1.915*	17.51***	16.99***
					0.273	0.092	0.000	0.000
IPCs in Sections A and G Only					2.730	4.103**	19.30***	18.72***
					0.108	0.021	0.000	0.000
IPCs in Sections A, C and G Only					0.958	1.817	16.74***	16.02***
					0.469	0.150	0.000	0.000
Lead Inventor Experience	0.014	0.015	-0.034	-0.052	-0.051	-0.045	-0.065	-0.100
	0.762	0.743	0.539	0.397	0.358	0.416	0.308	0.186
Number of Firm Patents		-0.001						-0.00803**
		0.581						0.020
Number of Firm Patents Squared		0.000						5.89e-06*
		0.514						0.074
Patent Generality				-1.044				
				0.310				
Technology Age in Years						-0.179***	-0.182**	-0.254**
						0.00484	0.042	0.024
Patent Priority Date in 1981-85							-1.687	-0.577
							0.395	0.832
Patent Priority Date in 1986-90							-1.366	-1.210
							0.392	0.534
Patent Priority Date in 1996-00							-3.448**	-3.653*
							0.0245	0.063
Constant	0.432	0.461	0.558	2.174	-2.183	-2.668	-17.56***	-17.02***
	0.342	0.349	0.434	0.276	0.386	0.297	0	0.000
Observations	134	134	134	108	133	133	114	114

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6b: Sample of firms that are older than 10 and younger than 20 years at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm technology age (i.e. time between the first patent that the firm filed and the agreement that it signed for the AMC patent) is larger than 10 and smaller than 20. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	6.1b	6.2b	6.3b	6.4b	6.5b	6.6b	6.7b	6.8b
Cosine Subclass Level	1.610**	1.787**	1.930**	2.634***	2.172**	3.022***	4.013***	4.397**
	0.021	0.046	0.013	0.002	0.010	0.002	0.003	0.027
Within Section Cosine, Group	-2.331***	-2.456**	-2.533***	-3.507***	-3.074***	-4.145***	-3.993***	-3.781***
	0.002	0.012	0.003	0.000	0.001	0.000	0.001	0.006
Patent Scope			-0.129	-0.158	0.023	0.087	0.542*	0.593
			0.333	0.364	0.901	0.624	0.092	0.118
Number of Patents Cited			0.039	0.053	0.031	0.024	0.0935*	0.111*
			0.159	0.138	0.425	0.585	0.056	0.075
Patent Cites Received Per Year			0.196	0.139	0.425	0.472*	0.404	0.482
			0.501	0.710	0.137	0.095	0.205	0.174
Patent Originality				-1.115	0.691	1.434	0.483	1.305
				0.502	0.733	0.505	0.815	0.571
Device					-0.541	-0.984	-2.839**	-2.752**
					0.459	0.210	0.017	0.046
IPCs in Section A Only					2.611**	3.501***	18.70***	17.98***
					0.032	0.005	0.000	0.000
IPCs in Section C Only					3.341***	3.901***	20.70***	20.42***
					0.010	0.003	0.000	0.000
IPCs in Sections A and C Only					1.288	1.915*	17.51***	16.99***
					0.273	0.092	0.000	0.000
IPCs in Sections A and G Only					2.730	4.103**	19.30***	18.72***
					0.108	0.021	0.000	0.000
IPCs in Sections A, C and G Only					0.958	1.817	16.74***	16.02***
					0.469	0.150	0.000	0.000
Lead Inventor Experience	0.014	0.015	-0.034	-0.052	-0.051	-0.045	-0.065	-0.100
	0.762	0.743	0.539	0.397	0.358	0.416	0.308	0.186
Number of Firm Patents		-0.001						-0.00803**
		0.581						0.020
Number of Firm Patents Squared		0.000						5.89e-06*
		0.514						0.074
Patent Generality				-1.044				
				0.310				
Technology Age in Years						-0.179***	-0.182**	-0.254**
						0.00484	0.042	0.024
Patent Priority Date in 1981-85							-1.687	-0.577
							0.395	0.832
Patent Priority Date in 1986-90							-1.366	-1.210
							0.392	0.534
Patent Priority Date in 1996-00							-3.448**	-3.653*
							0.0245	0.063
Constant	0.432	0.461	0.558	2.174	-2.183	-2.668	-17.56***	-17.02***
	0.342	0.349	0.434	0.276	0.386	0.297	0	0.000
Observations	134	134	134	108	133	133	114	114

*** p<0.01, ** p<0.05, * p<0.1

Table 1.7a Sample of firms that have more than 0 and less then 500 patents at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm has more than 0 and less then 500 patents. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients.

	7.1a	7.2a	7.3a	7.4a	7.5a	7.7a	7.7a	7.8a
Cosine Subclass Level	0.956*	0.915*	1.074**	2.043***	1.498**	1.466**	1.325*	1.455*
	0.080	0.091	0.048	0.008	0.028	0.034	0.081	0.075
Within Section Cosine, Group	-0.278	-0.349	-0.436	-1.598*	-0.766	-0.741	-0.801	-1.289
	0.646	0.563	0.483	0.058	0.320	0.341	0.377	0.199
Patent Scope			-0.176*	-0.167	-0.111	-0.114	0.334**	0.331**
			0.057	0.162	0.285	0.276	0.033	0.032
Number of Patents Cited			-0.014	0.005	-0.008	-0.007	-0.005	-0.008
			0.299	0.767	0.672	0.698	0.820	0.745
Patent Cites Received Per Year			0.470***	0.353*	0.585***	0.566***	0.272	0.413*
			0.008	0.089	0.000	0.001	0.158	0.059
Patent Originality				0.285	1.760**	1.683**	2.000**	1.806*
				0.746	0.019	0.027	0.028	0.070
Device					-0.386	-0.377	-0.514	-0.714
					0.333	0.343	0.301	0.175
IPCs in Section A Only					1.733*	1.707*	2.458**	2.233**
					0.058	0.062	0.015	0.037
IPCs in Section C Only					1.890*	1.892*	3.583***	3.574***
					0.057	0.055	0.002	0.002
IPCs in Section G Only					-1.329	-1.266	0.818	0.651
					0.395	0.413	0.602	0.684
IPCs in Sections A and C Only					1.149	1.138	2.556**	2.494**
					0.211	0.214	0.015	0.023
IPCs in Sections A and G Only					2.452**	2.377**	2.450*	2.987**
					0.033	0.041	0.091	0.041
IPCs in Sections A, C and G Only					-0.374	-0.425	-0.841	-0.915
					0.744	0.718	0.540	0.526
Technology Age in Years						0.018	-0.042	-0.053
						0.633	0.382	0.270
Patent Priority Date in 1981-85							1.241	1.310
							0.198	0.185
Patent Priority Date in 1986-90							1.485*	1.049
							0.065	0.189
Patent Priority Date in 1991-95							2.219***	2.398***
							0.003	0.001
Patent Priority Date in 1996-00							-1.132*	-1.310*
							0.078	0.052
Lead Inventor Experience	-0.007	-0.005	-0.016	-0.011	-0.013	-0.013	-0.003	0.002
	0.794	0.839	0.589	0.761	0.704	0.704	0.934	0.963
Number of Firm Patents		0.005						0.0147**
		0.225						0.011
Number of Firm Patents Squared		0.000						-4.67e-05***
		0.173						0.005
Patent Generality				-0.933				
				0.156				
Constant	-0.331	-0.396	0.038	0.292	-2.953**	-2.955**	-5.220***	-5.073***
	0.251	0.210	0.930	0.771	0.026	0.026	0.001	0.002
Observations	266	266	266	191	266	266	266	266

*** p<0.01, ** p<0.05, * p<0.1

Table 1.7b Sample of firms that have more than 500 patents at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm has more than 500 patents. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	7.1b	7.2b	7.3b	7.4b	7.5b	7.7b	7.7b	7.8b
Cosine Subclass Level	1.510	1.683	0.335	0.315	3.400*	0.493	1.803	7.380***
	0.237	0.159	0.819	0.866	0.061	0.812	0.365	0.008
Within Section Cosine, Group	-2.406***	-1.996***	-2.605***	-4.194***	-4.014***	-2.920**	-2.787**	-3.484***
	0.000	0.003	0.001	0.001	0.000	0.015	0.024	0.004
Patent Scope			0.171**	0.159*	0.056	0.057	0.257	0.475**
			0.013	0.050	0.511	0.582	0.121	0.047
Number of Patents Cited			0.006	0.017	0.0619***	0.0443*	0.016	0.034
			0.661	0.453	0.007	0.078	0.532	0.343
Patent Cites Received Per Year			0.667**	0.868*	1.367***	1.257***	1.113***	1.446***
			0.025	0.053	0.003	0.001	0.005	0.002
Patent Originality				1.469	2.614	2.615	1.548	3.103
				0.387	0.107	0.128	0.464	0.311
Device					-3.683***	-3.017**	-1.310	-1.775
					0.004	0.018	0.317	0.286
IPCs in Section A Only					13.60***	13.53***	9.123***	9.607***
					0.000	0.000	0.000	0.000
IPCs in Section C Only					16.44***	15.46***	12.29***	14.25***
					0.000	0.000	0.000	0.000
IPCs in Section G Only					18.15***	17.64***	12.49***	13.81***
					0.000	0.000	0.000	0.000
IPCs in Sections A and C Only					14.38***	14.10***	10.46***	10.48***
					0.000	0.000	0.000	0.000
IPCs in Sections A and G Only					16.85***	16.95***	13.47***	15.03***
					0.000	0.000	0.000	0.000
IPCs in Sections A, C and G Only					14.57***	14.51***	11.17***	12.00***
					0.000	0.000	0.000	0.000
Technology Age in Years						0.247***	0.260***	0.275**
						0.000	0.007	0.013
Patent Priority Date in 1981-85							4.600**	7.769***
							0.048	0.000
Patent Priority Date in 1986-90							3.913*	5.313***
							0.054	0.005
Patent Priority Date in 1991-95							1.843	2.079
							0.366	0.270
Patent Priority Date in 1996-00							1.495	1.150
							0.470	0.556
Lead Inventor Experience	-0.004	-0.009	0.022	0.017	-0.032	0.008	0.033	0.040
	0.918	0.815	0.614	0.768	0.601	0.895	0.691	0.659
Number of Firm Patents		0.000						0.00108***
		0.321						0.001
Number of Firm Patents Squared		0.000						-5.66e-08**
		0.689						0.003
Patent Generality				-1.282				
				0.198				
Constant	0.797*	0.066	-0.166	-0.210	-16.71***	-17.56***	-15.89***	-23.23***
	0.069	0.910	0.755	0.890	0.000	0.000	0.000	0.000
Observations	158	158	158	113	158	158	158	158

*** p<0.01, ** p<0.05, * p<0.1

Table 1.8 First Agreement Models

This table shows logit models with a dependent variable equal to 1 if the FIRST confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair. **Only the first agreement for each patent was selected.** Patent level measures correspond to the hospital patent which is under the agreement. **Firms for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0 in these models**. P-values under coefficients. Robust Standard Errors

	8.1	8.2	8.3	8.4	8.5	8.7	8.7	8.8
Cosine Subclass Level	2.095**	1.465*	1.664*	3.993***	1.847*	1.722	2.725**	2.744**
	0.017	0.090	0.065	0.002	0.090	0.114	0.021	0.021
Within Section Cosine, Group	-2.651***	-1.856**	-2.325**	-4.984***	-2.507**	-2.519**	-2.473**	-2.460**
	0.002	0.033	0.011	0.001	0.029	0.022	0.018	0.031
Patent Scope			0.101**	-0.003	0.145*	0.106	0.438**	0.448***
			0.042	0.971	0.051	0.252	0.011	0.009
Number of Patents Cited			-0.001	-0.012	0.011	0.008	0.010	0.012
			0.943	0.608	0.540	0.618	0.731	0.660
Patent Cites Received Per Year			0.299	0.062	0.344	0.350	-0.137	-0.134
			0.186	0.531	0.128	0.113	0.109	0.110
Patent Originality				0.207	0.964	0.935	0.953	1.159
				0.861	0.252	0.287	0.494	0.469
Device					0.097	0.031	0.343	0.380
					0.857	0.957	0.645	0.603
IPCs in Section A Only					2.117**	2.331**	3.986**	3.803**
					0.020	0.031	0.010	0.013
IPCs in Section C Only					3.147***	3.164***	6.434***	6.290***
					0.002	0.006	0.003	0.002
IPCs in Section G Only					1.287	1.287	4.725**	4.085**
					0.258	0.311	0.016	0.022
IPCs in Sections A and C Only					2.074**	2.057*	4.641**	4.441**
					0.022	0.053	0.015	0.018
IPCs in Sections A and G Only					1.924	1.993	3.914	4.154
					0.135	0.154	0.136	0.146
IPCs in Sections A, C and G Only					2.584**	2.772**	4.991**	6.382*
					0.028	0.035	0.023	0.054
Technology Age in Years						0.161**	-0.058	-0.132
						0.021	0.588	0.410
Patent Priority Date in 1981-85							3.338**	3.556**
							0.027	0.039
Patent Priority Date in 1986-90							5.322***	5.178***
							0.000	0.000
Patent Priority Date in 1991-95							4.809***	4.518***
							0.000	0.001
Patent Priority Date in 1996-00							-0.249	-0.300
							0.788	0.760
Lead Inventor Experience			-0.026	-0.0989**	-0.024	-0.024	-0.065	-0.064
			0.436	0.038	0.518	0.528	0.233	0.258
Number of Firm Patents		-0.000661**						-0.001
		0.010						0.306
Number of Firm Patents Squared		6.67e-08***						0.000
		0.004						0.272
Patent Generality				-3.808***				
				0.003				
Constant	0.576***	0.515**	0.267	4.183**	-2.811**	-3.138**	-7.631***	-7.324**
	0.006	0.022	0.480	0.014	0.026	0.035	0.010	0.014
Observations	194	194	194	145	194	194	194	194

*** p<0.01, ** p<0.05, * p<0.1