HHS Public Access

Author manuscript

Annu Rev Econom. Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

Annu Rev Econom. 2017 August ; 9: 441–469. doi:10.1146/annurev-economics-110216-100959.

How do patents affect research investments?*

Heidi L. Williams

MIT and NBER

1 Introduction

Technological change is widely perceived to be a key driver of improved standards of living. As an example, consider the gains in longevity that have been realized in the United States over the last half-century.[1] There is a widespread consensus among health economists – albeit, based on somewhat indirect evidence – that the realized improvements in longevity over this period largely reflect benefits reaped from technological change in medicine. For example, David Cutler's book *Your Money or Your Life* summarizes evidence on this point from a number of case studies which suggest that medical innovation has dramatically improved the value of treatments for at-risk infants, mental health patients, and heart attack patients over this time period (Cutler, 2004).

Although many factors - such as breakthroughs in basic science – influence the pace and direction of technological change, at least since the seminal work of Schmookler (1966) economists have focused on the important role of market incentives in shaping investments in new technologies. Because it is often difficult for inventors to capture the full social value of their inventions, it has long been recognized that – in the absence of government intervention – competitive private markets may provide less innovation than is socially desirable (Nelson, 1959; Arrow, 1962).[2] This concern has motivated long-standing academic and policy interest in the design of public policies to increase incentives for innovation.

The patent system is one of a suite of policy levers that has been used to attempt to bring the private returns captured by inventors closer to the social value of their inventions. By providing inventors with a temporary period of market power, patents aim to allow inventors to recoup the fixed costs of their research investments. The goal is for this temporary period of supra-competitive pricing to increase research investments relative to what would be realized in a competitive market with no government intervention. In addition, the patent system requires disclosure of the invention in exchange for the patent right. While the

*I am extremely grateful to Tim Bresnahan for very thoughtful and constructive comments, and to Ivan Badinski and Myles Wagner for excellent research assistance. Research reported in this publication was supported by the National Institute on Aging and the NIH Common Fund, Office of the NIH Director, through Grant U01-AG046708 to the National Bureau of Economic Research (NBER); the content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NBER. Financial support from NSF Grant Number 1151497 is also gratefully acknowledged.

Contact: heidiw@mit.edu.

[1]Nordhaus (2003) estimates that growth in longevity in the United States since 1950 has been as valuable as growth in all other forms of consumption combined.

[2]Of course, there could also be too much innovation due to business stealing effects. Recent empirical work by Bloom, Schankerman and Van Reenen (2013) suggests that on net (given current policy interventions) technological spillovers outweigh such business stealing effects.

disclosure requirement may decrease the private benefit of patenting at least in some markets, the motivation of the disclosure requirement is to increase the social value generated by the patent system.

While patent systems have been quite widely used both historically and internationally, there is nonetheless a tremendous amount of controversy over whether the patent system is – in practice – improving the alignment between private returns and social contributions. In 1951, Edith Penrose famously wrote, "[i]f national patent laws did not exist, it would be difficult to make a conclusive case for introducing them; but the fact that they do exist shifts the burden of proof and it is equally difficult to make a really conclusive case for abolishing them" (Penrose, 1951).[3] Over a half-century later, the question of whether patents are effective in spurring the development of new socially valuable technologies is still at the center of an active debate. On one hand, in a recent *Journal of Economic Perspectives* article Boldrin and Levine (2013) argue that the case against patents can be summarized concisely: "…there is no empirical evidence that [patents] serve to increase innovation…"[4] On the other hand, in a recent *George Mason Law Review* article Haber (2016) argues that "…the weight of the evidence supports the claim of a positive causal relationship between the strength of patent rights and innovation."

In this paper, I argue that estimates of three key parameters are needed to inform the question of how patents affect research investments.[5] First, how does the disclosure function of the patent system affect research investments? Second, to what extent is stronger patent protection (that is, longer patent terms or broader patent scope) effective in inducing additional research investments? For example, if we had longer patent terms or broader patent rights, would that spur the development of more pharmaceutical treatments? And third, do patents on existing technologies affect subsequent research investments? For example, do patent rights covering human genes affect the likelihood that gene-based pharmaceutical treatments and diagnostic tests are developed that are based on those genes? While answers to these three questions are not sufficient for informing optimal patent policy design, estimates of these three parameters would characterize how patents affect research investments, which is one important input into optimal policy design. For example, if patent-granted exclusivity rights induce socially valuable research investments, then there is a stronger case for longer or broader patents. On the other hand, if patents hinder subsequent innovation, then there is a weaker case.

Large literatures in economics, finance, management, strategy, law, and related fields have developed over the past few decades to investigate various aspects of the patent system. To name but a few examples, researchers have investigated what role patents play in encouraging the international diffusion of new pharmaceutical drugs (Cockburn, Lanjouw

[3]Machlup (1958) made a similar argument: "No economist, on the basis of present knowledge, could possibly state with certainty that the patent system, as it now operates, confers a net benefit or a net loss upon society. The best he can do is to state assumptions and make guesses about the extent to which reality corresponds to these assumptions…If we did not have a patent system, it would be irresponsible, on the basis of our present knowledge of its economic consequences, to recommend instituting one. But since we have had a patent system for a long time, it would be irresponsible, on the basis of our present knowledge, to recommend abolishing it."
[4]Plant (1934) famously made an earlier, similar argument: "…the science of economics as it stands to-day furnishes no basis of justification for this enormous experiment in the encouragement of a particular activity by enabling monopolistic price control."
[5]Of course, other parameter estimates would be needed to inform broader questions, such as the questions of optimal patent length and breadth.

and Schankerman, 2016; Duggan, Garthwaite and Goyal, 2016; Kyle and Qian, 2014), the role of patents in enabling technology licensing (Gans, Hsu and Stern, 2008), how to set royalties for standard-essential patents (Lemley and Shapiro, 2013), and whether the establishment of the US Court of Appeals of the Federal Circuit caused growth in patenting (Kortum and Lerner, 1999). The well-known paper by Hall, Jaffe and Trajtenberg (2001) documenting a linkage between the Compustat data and granted US patents has almost 3,000 citations (Google Scholar, August 6 2016), one summary indicator of the volume of patent-related research using that data set alone. However, somewhat surprisingly, there has been remarkably little empirical research tackling the three research questions described above, which I argue are the key empirical parameters that we need in order to confidently make statements about whether patents are effective in encouraging the development of new technologies.

The difficulty inherent in providing answers to the three questions outlined above is that answering these questions requires constructing counterfactuals within which we can measure the set of "potential" innovations – that is, innovations that could have been developed in the sense of being scientifically feasible, but that were never brought to market. If we measure the set of potential inventions and can compare that with the observed set of commercialized inventions, then we can quantify the existence of – and, ideally, the social value of – the "missing" technologies that could have been developed had we had a better-designed patent system.

Section 2 provides a brief primer on patents. I then describe in more detail these three parameters – how the disclosure function affects research investments, how patent strength affects research investments in new technologies, and how patents on existing technologies affect follow-on innovation – and review the available evidence which has attempted to empirically estimate these parameters (Sections 3, 4 and 5). Section 6 concludes.

2 Patents: A brief primer

In the US, inventors wishing to obtain a patent submit an application to the US Patent and Trademark Office (USPTO).[6] Patent applications are comprised of two key sections. First, the "specification" is a written description of the invention which includes references to so-called "prior art," which are citations to previously filed patent applications, previously granted patents, prior scientific publications, and other sources which are known to the inventor and relevant to the patentability of the invention. Second, the "claims" of the patent are a specific list of what the applicant wishes to claim intellectual property over. Filing and other fees must be paid by applicants.[7]

The USPTO is responsible for determining whether inventions claimed in patent applications qualify for patentability. The standard for patentability is that inventions are patent-eligible (35 U.S.C. §101), novel (35 U.S.C. §102), non-obvious (35 U.S.C. §103),

[6]Patent protection is jurisdiction-specific: if an inventor wants to hold patent protection in each of the US, France, and Japan, she must file her patent application in each jurisdiction. As described by Putnam (1996), the timing of international filings relative to an initial filing is constrained under the Paris Convention.

[7]For more information on the current USPTO fee schedule, see: https://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule.

useful (35 U.S.C. §101), and that the text of the application satisfies the disclosure requirement (35 U.S.C. §112). An initial decision on whether a given patent application meets this standard for patentability is made by a patent examiner with expertise in the scientific or technical area of the application. If the examiner issues an initial allowance of the application, the inventor can then be granted a patent. In most cases, the examiner instead issues a so-called "rejection" as her first decision on the application. However, in practice patent applications cannot be rejected by the USPTO, only abandoned by applicants (Lemley and Sampat, 2008). What that means is that a rejection is essentially an invitation for the applicant to submit a revised patent application that, for example, eliminates one or more claims or changes the text of some claims. The patent prosecution process – a phrase used to refer to the interaction between the USPTO and the patent applicant or her representative (such as a lawyer) – can involve several rounds of "rejection" and revision, and in this sense can best be conceptualized as an iterative process between the applicant and the examiner, rather than as a one-time decision by the examiner. Applicants presumably choose between "revising and resubmitting" or abandoning a rejected patent application by weighing the relevant costs and benefits: if a revision that accommodates the examiner's criticisms would result in a patent that – even if granted – would be too narrow to provide much economic value to the applicant, we would expect the application to be abandoned. Hence, while much attention has focused on estimating what share of patent applications are allowed by the USPTO, attention on that metric is somewhat misplaced: presumably almost any patent application would be allowed if the applicant were willing to accept a sufficiently narrowed set of claims. However, claims that are very narrow would provide very little economic value, and hence applicants will generally not find it in their private interest to prosecute such applications through to the allowance stage.[8]

The institutional structure of the patent system could affect incentives for innovation in a number of ways, but theoretical models of optimal patent design have largely focused on two parameters: patent term length and patent breadth. The US patent term length is currently 20 years from the filing date of the patent.[9] While there exists some cross-country variation in the patent term, this 20 year term exists in most other countries as well (largely due to international harmonization efforts such as the World Trade Organization's Agreement on Trade-Related Aspects of Intellectual Property Rights). As we discuss below, a large literature starting with Nordhaus (1969) has investigated how patent term length might impact incentives for innovation.

Patent breadth (or scope) is less straightforward to define. From a theoretical perspective, the economic meaning of patent breadth is clear: how different must rival products be in order to

[8]USPTO "rejection rates" are also difficult to measure for other reasons, including the fact that US patent applications can spawn closely related "new" applications (such as continuations or divisionals). Carley, Hegde and Marco (forthcoming) use internal USPTO data to calculate simple and "family" (including patents granted to continuations or divisionals) grant rates in the universe of new (not related to any previously filed applications) utility patent applications filed at the USPTO from 1996–2005. In that sample, 55.8% of applications were granted patents directly, and including patents granted to children increases the allowance rate to 71.2%. See also Lemley and Sampat (2008).

[9]This is true for utility patent applications filed on or after 8 June 1995. Technically, the 20 year term runs from the filing date of the earliest US filing or the earliest patent cooperation treaty (PCT) international filing to which priority is claimed, excluding provisional applications (35 U.S.C. §154). The only exception to the 20 year term of which I am aware within the US is the so-called Hatch-Waxman term extension provided to some pharmaceutical innovations (35 U.S.C. §156), which compensates for delays accruing due to regulatory approval by other US agencies.

be deemed non-infringing on a given patented product (see, e.g. Gilbert and Shapiro (1990) and Klemperer (1990))? But from an empirical perspective, measuring the breadth of patent applications or granted patents is quite challenging. Most definitions are based in some way on the text of the claims of the patent, which can be conceptualized as the "boundaries" of the intellectual property right being sought or having been granted. Marco, Sarnoff and DeGrazia (2016) empirically investigate two of the more commonly-used metrics for patent scope: the number of words in the claims, and the total number of claims.[10] As shown in their analysis, the patent prosecution process tends to narrow the scope of patent claims as measured by both claim length and claim count. As motivated by the discussion above, "rejections" (abandonments) of patent applications can arguably be considered one extreme on the continuum of claim narrowing that tends to occur during the patent prosecution process. To the extent that inventors (or potential inventors) correctly anticipate this claim narrowing, this is a second mechanism through which the institutional structure of the patent system could affect incentives for innovation.

Once granted, in order to keep a patent in force the owner must pay maintenance fees. For the USPTO, these fees are currently due at 3.5 years ($1,600), 7.5 years ($3,600), and 11.5 years ($7,400). Pakes (1986) and Schankerman and Pakes (1986) pioneered the idea of using renewal fees to provide lower-bound estimates on the private value of granted patents.

3 The disclosure function of the patent system

Part of the 'grand bargain' struck by the design of the patent system is that in exchange for receiving a patent, inventors must publicly disclose their invention. This is in contrast with keeping their invention secret, which would be done – for example – with trade secrecy. The so-called disclosure theory of patents stresses that a major benefit of the patent system is that these disclosures will stimulate the development of new ideas. In practice, very little evidence is available on whether or how the disclosure function of the patent system affects research investments in practice.

One way in which the disclosure function of the patent system could affect research investments is if, by making patenting less attractive to private inventors, the disclosure function reduces the level of incentives provided by the patent system. Firms report in some surveys that they choose not to patent some inventions because of the disclosure requirement (Cohen et al. 2000; Cohen et al. 2002; Levin et al. 1987); whether this in turn reduces R&D incentives is of course an empirical question.

Setting aside this incentive effect, the disclosure theory of patents focuses on an enablement effect: that is, conditional on an invention being disclosed, this theory asserts that new research ideas will then be stimulated. Critics of the disclosure theory of patents have argued that disclosure may generate few social benefits in practice if, for example, actual patents contain little valuable technical information, or if only inventions that would be disclosed

[10]One important limitation of these types of measures of patent breadth is that in practice breadth is determined not only by the text of patent, but also by courts' interpretations of these claims, and the market's expectations of how the courts will interpret these claims. Because only a subset of patents are litigated, and because market expectations are difficult to measure, it is challenging to construct measures of patent breadth that incorporate those factors.

anyways are patented (Roin, 2005). These are all of course again empirical questions. Ouellette (2012) develops some empirical evidence on these points by analyzing a survey of nanotechnology researchers. She documents that 64% of survey respondents report having read patents, of which 70% used them as a source of technical information.

A natural setting in which to attempt to analyze the effects of disclosure requirements on research investments would be to analyze a recent set of policy changes in the US and elsewhere that have expanded the scope of disclosure by requiring disclosure not just of granted patents, but rather of all patent applications (with some opportunities to 'opt-out' of disclosure for certain types of patent applications which are not granted patents). I am aware of two papers which have analyzed this policy variation in the US, although neither has aimed to quantify the effects of this policy change on research investments. First, Graham and Hegde (2015) analyze the 'opt-out' decision among eligible applicants (namely, applicants not seeking foreign patent protection), and document that 85% of these applicants opt for disclosure. They argue these results suggest that, contrary to common belief, there may be private benefits to disclosure. Second, Hegde and Luo (forthcoming) analyze the effect of patent disclosure on patent licensing in the biomedical industry. They document that after the passage of the American Inventors Protection Act (AIPA), which mandated publication of patent applications 18 months after filing, US patent applications were 20 percent less likely to be licensed after grant and significantly more likely to be licensed between publication and grant. They argue that these results suggest that disclosure facilitates sales and transactions in the market for ideas. I hope that future researchers will navigate a way to use this AIPA-type policy variation in the US or other countries to examine the question of how disclosure affects research investments, both through the incentive channel and through the enablement effect.[11]

4 Patents and research investments in a Nordhaus (1969) framework

To the best of my knowledge, Nordhaus (1969) provided the first formal theoretical framework analyzing optimal patent length.[12] In Section 4.1, I summarize Budish, Roin and Williams (2016)'s simplified and slightly modified version of the Nordhaus model, and in Section 4.2 I discuss survey evidence as well as empirical papers which have attempted to estimate the key empirical elasticity which emerges from this framework.

4.1 Theory

In the original Nordhaus (1969) model, a representative firm chooses its level of R&D investment as a scalar decision variable, and R&D generates both private and social benefits by lowering that firm's production costs for a single output good. In Budish, Roin and Williams (2016), in order to focus attention on empirical elasticities, we instead model a representative firm's decision over which potential R&D projects to invest in, and we

[11]Both of the studies discussed in Section 5.2 on how patents affect follow-on innovation hold disclosure constant: in the case of Galasso and Schankerman (2015), invalidated patents have already been disclosed; in the case of Sampat and Williams (2015), both accepted and rejected patent applications are disclosed during the time period of their sample.

[12]In particular, I mean to refer to Nordhaus (1969) Chapter 5, pages 76–86.

assume that R&D investment generates private and social benefits by bringing to market inventions that would not have otherwise existed.

The set of potential inventions is of unit mass; index these potential inventions by $i \in I$. The R&D investment associated with pursing a potential invention i has a cost c_i and a probability of success p_i. If successful, the invention will generate social value for T_i years, after which it will become technologically obsolete. Patents can be conceptualized in this stylized model as providing a fixed number of years t_{patent} during which a successful invention can be sold by a monopolist, after which there can be free entry. While successful inventions are non-obsolete and are priced by a monopolist, they generate annual profits of π_i and annual social value of v_i^m. While successful inventions are non-obsolete and priced competitively, they generate annual social value of v_i^c. We ignore discounting by assuming that R&D investment is a one-time cost incurred at time 0, and assuming that the annual quantities π_i, v_i^m, and v_i^c grow at a discount rate which is the same for the firm and society.

The firm will choose to invest R&D in potential invention i if and only if the number of years of expected monopoly life (that is, the minimum of t_{patent} and the socially useful life of the technology T_i) times the risk-adjusted profits per year (p_i times π_i) exceeds the cost of the R&D investment c_i: $p_i \cdot \min(t_{patent}, T_i) \cdot \pi_i \geq c_i$. The social value of the firm bringing invention i to market is $p_i \cdot [\min(t_{patent}, T_i) \cdot v_i^m + (T_i - \min(t_{patent}, T_i)) \cdot v_i^c] - c_i$.

Budish, Roin and Williams (2016) use this simple framework to derive expressions for the benefits and costs of a marginal increase in the patent term t_{patent}. The benefit of a marginal increase in the patent term is that more inventions will be elicited on the margin. Letting ξ denote the quantity of inventions elicited at the margin, we can write these benefits as:

$$\text{Benefits} = \underbrace{\xi}_{\substack{\text{elasticity of R\&D} \\ \text{wrt patent term}}} \times \underbrace{\mathbb{E}_{EML_i \cdot \pi_i = c_i}[ETL_i \cdot v_i^c - EML_i \cdot (v_i^c - v_i^m) - c_i]}_{\text{social value of marginal inventions}}. \quad (1)$$

The cost of a marginal increase in the patent term is that inframarginal inventions – that is, inventions that would have been developed even in the absence of the patent term extension – spend a larger share of their socially useful life under monopoly pricing, which generates additional deadweight loss. We can write these costs as:

$$\text{Costs} = \int_I \underbrace{\mathbf{1}_{\{EML_i \cdot \pi_i \geq c_i\}} \mathbf{1}_{\{T_i > t_{patent}\}}}_{\text{intensive margin}} \times \underbrace{(v_i^c - v_i^m)}_{\text{deadweight loss}} \, di. \quad (2)$$

Optimal policy in the Nordhaus (1969) framework equates these benefits and costs at the margin. The ξ parameter in Equation (1) quantifies the extent to which stronger patent protection (in the form of longer patent terms) are effective in inducing additional research investments. As we discuss in Section 4.2, while it is straightforward to define this parameter theoretically, this parameter is difficult to empirically estimate because it requires drawing inferences about inventions that could have been developed – in the sense of being

scientifically feasible – but which were never brought to market because the prevailing patent term was too short to incentivize the necessary private research investments.

4.2 Empirics

4.2.1 Survey evidence—Three influential surveys have documented evidence on the importance of patents in spurring research investments: Mansfield (1986), the so-called Yale survey (Levin, Klevorick, Nelson and Winter, 1987), and the so-called Carnegie Mellon survey (Cohen, Nelson and Walsh, 2000).

Mansfield (1986) conducted a survey aimed at shedding light on the impact of patents on the rate of development of and commercialization of new inventions. The survey included a random sample of 100 firms from twelve US industries.[13] The survey asked respondents to estimate the proportion of its inventions developed in 1981–83 which would not have been developed had the firm been unable to obtain patent protection.[14] Table 1 summarizes the key tabulation of his results, across industries. In seven of the twelve industries surveyed, patent protection is estimated to be of limited importance in both the development and commercialization of new inventions. In only two industries - chemicals and pharmaceuticals - are patents deemed to be essential to the introduction of 30 percent or more of inventions. Table 2 of Mansfield (1986) (not shown here) documents that in industries where patents are self-reported to be more important (as in Table 1), firms also self-report patenting a higher share of patentable inventions.

Levin, Klevorick, Nelson and Winter (1987) report tabulations from the so-called Yale survey which aimed to shed light on the R&D appropriability conditions in different industries. The authors identified a sample of publicly traded firms with R&D expenses, and sent questionnaires to employees (determined through communication with the CEO) with knowledge of the major lines of business at the firm.[15] The authors asked respondents to rate the effectiveness of alternative means of capturing and protecting the competitive advantages of new or improved processes and products on a seven-point scale: patents to prevent duplication, patents to secure royalty income, secrecy, lead time, moving quickly down the learning curve, and sales or service efforts. In aggregate, patents were reported to be the least effective mechanism of appropriation for new processes, with secrecy being the most effective. Product patents were reported to be relatively more effective, although less so than several other strategies such as lead time and sales or service efforts. Table 2 documents tabulations of these survey responses at the industry level, for the eighteen industries with ten or more survey responses. As in Mansfield (1986), the chemical industries stand out as reporting higher patent effectiveness, for both process and product patents. Even among that

[13]The sampling frame was all firms on a list published in *Business Week* in 1982 of firms in those twelve industries spending over $1 million (or 1 percent of sales, if sales were at least $35 million) on R&D in 1981, as well as a group of smaller firms listed in *Poor's Register*. He notes that "practically all" surveyed firms filled out a detailed questionnaire, with only 4 of 100 firms not responding, and the major executives of 25 firms were also interviewed.

[14]An "invention" was defined as a prescription for a new or improved product or process (or one or more components or facets thereof) that is prospectively useful and that is not obvious to one skilled in the relevant art at the time the idea is generated.

[15]The sampling frame was again the *Business Week* annual R&D survey, which was used to identify all publicly traded firms that reported R&D expenses in excess of 1% of sales or $35 (a total of 746 firms in 1981). Dun and Bradstreet's *Million Dollar Directory* was used to assign each firm to its major line of business. The authors received 650 responses (representing 130 lines of business) from the final sampling frame of 1,562 units.

group, only the pharmaceutical (drugs) industry reported patents to be the most effective mechanism of appropriation for both product and process patents. Levin, Klevorick, Nelson and Winter (1987) conjecture that patents may be more effective in the chemical industry due to relatively clearer standards of patent validity and/or infringement due to molecules being discrete and differentiable.

Cohen, Nelson and Walsh (2000) report tabulations from the so-called Carnegie Mellon survey, which built closely on the Yale survey and was aimed at evaluating whether appropriability conditions changed over time with a series of reforms in the 1980s which were perceived as strengthening the patent system. The authors surveyed R&D labs/units of manufacturing firms.[16] The survey asked respondents about the effectiveness of various appropriability mechanisms (patents, secrecy, lead time, complementary sales and service, and complementary manufacturing facilities and know-how), where effectiveness was measured as the percentage of product or process innovations for which each appropriability mechanism was effective (less than 10%, 10–40%, 41–60%, 61–90%, greater than 90%). Tables 1 and 2 of Cohen, Nelson and Walsh (2000) (not shown) indicate, consistent with the earlier Mansfield (1986) and Yale surveys, that patents are reported to be important for product innovations in the drug and medical equipment industries, but in no industry are patents regarded as the most effective appropriability mechanisms.

Across industries, all three surveys documented evidence that on average firms report that patents have limited effectiveness as an appropriation mechanism relative to other levers such as lead time and trade secrecy, but that in a few industries – namely, chemicals and pharmaceuticals – firms report that patents are essential for spurring R&D investments. This survey evidence is useful to keep in mind partly because several of the empirical analyses of the relationship between patents and research investments described below focus on particular industries.

4.2.2 Evidence from patent law changes—Moving beyond survey evidence to empirically estimate the relationship between patent strength and research investments requires finding variation in the patent strength that we (as a society) provide to different types of potential inventions. Several researchers have investigated using changes in patent laws as a source of such variation.

Sakakibara and Branstetter (2001) is the first such analysis of which I am aware, providing a very careful study of a 1988 Japanese patent reform. While this reform packaged a number of legal changes, the authors argue that perhaps the most important feature was a change from a single-claim patent system to the type of multi-claim patent system common in the US and elsewhere, in which one patent application can include multiple claims. They note that patent experts argue that this reform broadened the scope of patent protection provided under Japanese law. Using data from the Japanese Patent Office, they document a sharp and immediate increase in the number of claims per patent application, consistent with this

[16]The sampling frame of 3,240 labs was randomly drawn from eligible labs in the Directory of American Research and Technology or belonging to firms in Standard and Poor's Compustat database stratified by 3-digit SIC code. The authors received 1,478 responses and further restricted the sample to firms with at least $5 million in sales or business units with at least 20 employees for a final sample of 1,165 observations.

reform being salient to applicants. The authors then use data on the R&D investments of around 300 publicly traded Japanese manufacturing firms to empirically test whether broader patent rights – as induced by this reform – increase research investments. The time series evidence presented in Figure 1 (which reproduces Figure 5 of their paper) documents that while there was an increase over time in R&D investments that predates the 1988 reform, there is little evidence for a deviation in this trend after the 1988 reform. Based on this as well as similar evidence from a variety of similar analyses, the authors conclude that while the patent reform clearly induced a behavioral response by Japanese firms (as evidence by the change in claim behavior), there is little evidence that the broader patent protection provided by the reform resulted in increased levels of R&D investments.

Perhaps the best-known analysis of patent law changes is that of Lerner (2009). Lerner collected records of major patent policy changes in 60 countries over 150 years. Because cross-country data on R&D investments analogous to that used by Sakakibara and Branstetter (2001) is not available historically, Lerner instead constructs a novel measure of R&D investment: he chooses a country (Great Britain) that had a relatively stable patent policy over the time period of study, and analyzes how foreigners' decisions to file patents in Great Britain change as a function of patent law changes in their home country. For example, if Japan lengthens its patent term, do Japanese firms increase their propensity to apply for patents in Great Britain? The advantage of this approach is that even though Japan's patent law change may induce changes in firms' decisions over whether or not to file for patent protection in Japan on existing research investments, filings by Japanese firms in Great Britain should change only if these Japanese firms change their research investments. Figure 2 (which reproduces Figure 2 of Lerner (2009)) graphically depicts the average change in patent applications around the date of patent-strengthening policy changes, net of an index that controls for changes in patent filings that are common across countries within a year. Filings by domestic entities in either their home country (thin solid line) or in Great Britain (dashed line) were flat or if anything fell after the patent laws were strengthened. Based on this as well as similar evidence from a variety of similar analyses, Lerner concludes that there is little evidence that stronger patent laws result in increases in R&D investments as measured by patent filings.

As noted by the authors of these two papers, the lack of an observed relationship between stronger patent rights and research investments is a puzzling result. Lerner (2009), for example, notes: "It runs not only against our intuition as economists that incentives affect behavior, but also counter to the findings in the 'law and finance' literature that stronger property rights…encourage economic growth." In other work joint with Eric Budish and Ben Roin (Budish, Roin and Williams, 2016), I have argued that these results are less puzzling than they may seem on first glance. As noted by the authors of these studies, as implemented in practice – often as part of trade negotiations – patent law changes may affect domestic R&D through other mechanisms such as changes in foreign competition. More generally, these types of studies by construction investigate how R&D investments by domestic firms respond to domestic changes in patent laws, but it isn't clear that this thought experiment is the right one. First, if a "large" economy (like the US) were to lengthen its patent term, that would affect the investment incentives not only of US firms but also of non-US firms that aim to sell their goods to US consumers. Second, because technologies are

developed for a global market, patent law changes in "small" economies likely generate relatively small sources of variation in global R&D incentives. For both reasons, we might expect estimates of the elasticity of research investments with respect to patent terms that are based on country-specific law changes to be biased towards zero.

4.2.3 Evidence from variation in clinical trial length—Given the limitations inherent in using patent law changes as variation, a natural question is whether we can isolate other alternative forms of variation in effective patent terms. Budish, Roin and Williams (2015), described below, provide one such attempt.

On paper, the US patent system provides an identical 20 year patent term to all potential inventions. However, in practice some inventions receive fewer than 20 years of effective patent protection. One example of such uneven effective patent protection arises in the case of drug development. Before a drug can be sold to consumers, firms are required by regulators such as the US Food and Drug Administration (FDA) to document evidence from randomized clinical trials that their drug is safe and effective. However, private firms face strong economic and legal incentives to file for patent protection on candidate drug compounds very early in the drug discovery process, *prior* to starting clinical trials. What that means is that in practice, we – as a society – have designed our patent system to provide shorter effective patent terms to drugs developed to treat patient groups who require longer clinical trials: with a fixed patent term, every additional day in clinical trials is one fewer day that a firm can charge supra-competitive prices on a drug that is being sold to consumers.

Because evidence of "effectiveness" is traditionally defined by the FDA as evidence that a drug improves survival outcomes of patients, the length of a given clinical trial is in large part determined by the time it will take to show improved survival for the treated patient population. The simple statistics of power calculations imply that clinical trials will need to be longer if the drug targets a patient population with a lower mortality rate: for example, it will take less time to document a statistically significant improvement in survival among late-stage breast cancer patients than it will among early-stage breast cancer patients. Of course, having to conduct longer clinical trials may well matter for reasons other than receiving a shorter patent term: longer clinical trials likely cost more, both in direct financial costs and in a time value of money sense. Such alternative factors complicate use of this empirical context to directly estimate an elasticity of research investments with respect to the patent term, as we discuss more below.

In Budish, Roin and Williams (2015), we analyze this distortion in the context of cancer drug development. The treatment of cancer tends to be organized by the organ in which the cancer originates (e.g. breast, lung, prostate) and the stage of the cancer at the time of diagnosis (local, regional, metastatic). This medical categorization of cancer types provides a natural framework within which we attempt to categorize observed and potential research investments into cancer treatments. That is, even if we never observe private firms attempting to develop drugs to treatment patients with a particular type of cancer, we know from a medical perspective the full set of cancer patient groups that exist and can use this categorization to test for the existence of "missing inventions" due to a lack of economic incentives to make R&D investments in certain patient groups. The cancer-stage

classification outlined above is natural for this purpose in part because cancer drugs are approved by the FDA for treatment of a patient group at the same cancer stage level. For example, Genentech's drug Bevacizumab was approved by the FDA in 2004 for the treatment of patients with metastatic carcinoma of the colon and rectum. We construct measures of research investments for patients diagnosed with different cancer-stage types by parsing listings of which patient groups were eligible to enroll in different clinical trials, as these enrollment lists provide relatively precise information about which patient groups the sponsor was attempting to develop a new drug to treat.

For each cancer-stage, we also need to observe a proxy for how long it would likely take to develop a drug to treat that patient group. Because many cancer-stage patient groups have never had a drug developed, using observed clinical trial lengths is not possible.[17] We focus on survival time as a proxy for clinical trial length, given that survival time predicts potential clinical trial length in the statistical sense described above, and because for cancer patients relatively rich data is available tracking the survival time of cancer patients diagnosed with different cancer-stages (for example, from the SEER cancer registry).

We use this data to first document that – consistent with our conjectured distortion – patient groups requiring longer clinical trials (as proxied by having higher survival rates) tend to less frequently be the target of drug development efforts. Figure 3 provides one tabulation of this result. Metastatic cancer patients have a much lower five-year survival rate (on the order of 10%) than regional (on the order of 50%) or localized (on the order of 70%) cancer patients, and metastatic patients are much more frequently targeted by clinical trials – being allowed to enroll in around 12,000 clinical trials in our data, compared to around 10,000 for regional patients and around 6,000 for localized patients.

While consistent with our conjectured distortion, the pattern documented in Figure 3 is difficult to interpret for two reasons. First, a correlation between commercialization lags and research investments need not reflect a causal relationship. For example, it could be that the science behind developing drugs to treat localized cancers is just more difficult than the science behind developing drugs to treat metastatic cancers, so that even if clinical trials evaluating treatments for localized cancers were shorter we would not observe more of those trials. Second, even if this correlation did reflect a causal relationship, it may not be evidence of a distortion. As we clarify in a simple theoretical framework, both private firms and the social planner are more likely to pursue projects with shorter commercialization lags, which can be completed more quickly. Our paper presents the results of two empirical tests which together address these concerns.

First, we document evidence consistent with the idea that shortening clinical trial lengths can increase research investments. Figure 4 provides one tabulation of this result. Our key idea is to take advantage of the fact that for some cancers firms are not required to show that drug

[17]In addition to this non-observability constraint, using observed clinical trial lengths as a proxy for firms expectations would likely be biased, because the set of trials that have been conducted is – as posited by our theoretical model – not a random subset of the universe of potential trials. In particular, we would expect that clinical trials to treat patient groups with long survival times may be more likely to occur in practice if for some reason they are allowed to conduct shorter trials via, e.g. showing evidence that the drug improves a surrogate endpoint rather than improving survival.

treatments improve survival, but instead are allowed to show that drug treatments improve an intermediate or "surrogate" endpoint that is observed more quickly. As an example, clinical trial endpoints for leukemia treatments are generally based on blood cell counts and related bone marrow measures, rather than being based on observed survival outcomes. For cancers where such surrogate endpoints are consistently and predictably allowed, we expect to see less of a negative relationship between commercialization lags and research investments. Figure 4 contrasts hematologic cancers, which are essentially always approved on the basis of surrogate endpoints, with other cancers, and documents that among hematologic cancers there is not a negative relationship between five-year survival rates (our proxy for commercialization lags) and research investments.[18] This fact is consistent with there being a causal relationship where if firms are allowed to conduct shorter clinical trials for a given cancer, more trials will be conducted.

Second, we test for evidence of a distortion in private research investments by contrasting public and private research investments. While – consistent with our theoretical model – both private and public research investment levels are higher for clinical trials with shorter commercialization lags, the correlation between commercialization lags and research investments is economically and statistically significantly more negative for private research investments, consistent with our conjectured distortion.

Taken together, these results suggest that private research investments are distorted away from long-term research projects. As motivated above, the patent system is one factor that could be generating this distortion, but our analysis is not able to isolate the importance of patents per se as opposed to other factors. A back-of-the-envelope calculation based on the hematologic/non-hematologic comparison described above suggests a semi-elasticity of R&D investment with respect to a one-year change in commercialization lags between 7–23 percent. Only under the very strong assumption that commercialization lags affect R&D investments only through changing the patent term would that estimate correspond to the semi-elasticity of R&D investment with respect to the patent term.[19]

To summarize, evidence from patent law changes has provided little evidence that stronger patent rights encourage research investments, but such law changes are – I argue – likely under-powered to detect such effects. While the variation in clinical trial lengths analyzed by Budish, Roin and Williams (2015) provides evidence consistent with patent length having a quantitatively important impact on research investments, that study is not able to rule out several other plausible mechanisms for the results. Identifying new sources of variation and empirical contexts in which this relationship can be more convincingly estimated is a very important avenue for future research.

[18]The solid red linear fit line for hematologic cancers in fact displays an upward slope in Figure 4. Our theory gives no prediction on what the slope of this relationship should be, but rather only predicts the comparative static that the commercialization lag distortion should make the slope of this line more negative.

[19]Interpreting this estimate economically also requires making an assumption on whether factors like credit constraints or other input constraints are relevant, in the sense that having more research on certain types of projects could imply that we have less research on other types of projects. The pharmaceutical industry, as is the focus of this paper, is arguably not an industry where we think that credit constraints are relevant, but this issue is worth keeping in mind in interpreting the results of studies on this topic.

5 Patents and research investments in a Scotchmer (1991) framework

A relatively well-developed theoretical literature has explored the question of how patents on existing technologies will affect follow-on innovation. Consider as an example a patent on a human gene sequence, such as the gene patents granted to the firm Incyte Pharmaceuticals. Sequenced genetic data may be a useful input into follow-on scientific discoveries, such as the documentation of evidence that variations of a given gene are linked to various diseases, and into follow-on commercial applications of those scientific discoveries, such as the development of gene-based diagnostic tests or pharmaceutical treatments. The question in this case is whether patents on human genes will encourage or discourage the development of follow-on innovations such as gene-based diagnostic tests.

If Incyte had knowledge of all follow-on innovations related to their patented gene, private and social incentives would be aligned: Incyte has the private incentive to develop all socially valuable inventions. However, if ideas are "scarce" in the sense of Scotchmer (2004), some individuals or firms other than Incyte may have ideas for follow-on innovations that are not known to Incyte. In this case, follow-on inventions will require a license from Incyte. Green and Scotchmer (1995) clarify that if the "correct" licensing agreements are signed – namely, ex ante agreements which are signed prior to follow-on inventors sinking their research investments – then all socially valuable follow-on inventions will still be developed. However, if there are impediments to ex ante licensing agreements, such as asymmetric information, then follow-on innovation may be discouraged (e.g. Bessen (2004)).

Work dating back at least to Kitch (1977) has argued that patents may facilitate the coordination of investments in follow-on innovations, providing an incentive for maximizing the value of the original patent in a way that might not otherwise be realized if the original technology were instead in the public domain. This so-called "prospect theory" of patents provides one reason why we might expect a positive relationship between patents and follow-on innovation.

Taken together, this theoretical literature provides ambiguous predictions for how patent rights will impact follow-on innovation. In Section 5.1 I outline a simple theoretical framework aimed at clarifying the main ideas that have been discussed in the prior theoretical literature, with the goal of clarifying the key empirical elasticity which emerges from this literature.[20] In Section 5.2, I then discuss survey evidence as well as empirical papers which have attempted to estimate this elasticity.

5.1 Theory

Assume all actors are risk neutral.[21] One firm (Incyte) holds a set of upstream technologies (genes). There exists a supply of downstream product developers, each of whom has a

[20]Note that this simple framework fails to capture some important theoretical models in the prior literature, such as the contribution of Bessen and Maskin (2009), and also sets aside the disclosure function discussed in Section 3.

[21]This framework is drawn from unpublished work contained in my Harvard PhD dissertation; I am very grateful to Ed Glaeser for comments while I was a graduate student. This framework is very similar in spirit to the analytical framework presented in Galasso and Schankerman (2015).

probability π of discovering an innovation, which yields the ability to provide a downstream commercial application (gene-based diagnostic tests) at zero marginal cost. I assume that ideas are scarce and that the set of potential innovations is large enough to abstract away from cases in which two downstream product developers discover the same innovation (apart from generic entry, discussed below). The downstream market has demand $D(p)$, which I assume is differentiable and decreasing with respect to price (*i.e.* $D'(p) < 0$).

Ex post - that is, conditional on entry and successful product development - downstream product developers choose the optimal price to maximize profits:

$$max_p \left[\Pi = p \cdot D(p)\right] \quad (3)$$

Differentiating Equation (3) with respect to price p implies $p^m \cdot D'(p^m) + D(p^m) = 0$, which can be re-written to solve for the optimal monopoly price p^m as:

$$p^m = -\frac{D(p^m)}{D'(p^m)} \quad (4)$$

Let $q^m \equiv D(p^m)$ denote monopoly output. Then profits for the monopolist are:

$$\Pi^m = -\frac{D(p^m)^2}{D'(p^m)} \quad (5)$$

Each downstream product developer has an idiosyncratic opportunity cost of time, denoted c. Let the number of downstream product developers with costs below any given level x be denoted by the differentiable, strictly increasing function $N(x)$.

I consider two cases: first, when the upstream technologies are in the public domain; and second, when the upstream technologies are granted patents.

First, consider the case in which the upstream technologies are in the public domain, in which case any firm is free to independently develop downstream products.[22] In this case, assume there is a probability $1 - \delta$ that the downstream patent cannot be protected because other firms can costlessly create a generic product; in this case, the price of the downstream product is pushed to marginal cost (assumed to be zero) and the downstream product developer realizes zero profits. On the other hand, with probability δ the downstream patent can be protected, in which case the downstream firm earns monopoly profits. Firms will enter as long as their cost c is less than expected monopoly profits - adjusted by the probability of successful innovation π and the probability the downstream patent can be protected δ. Thus, firms will enter as long as c is less than $\pi \cdot \delta \cdot \Pi^m$, and the total amount of entry will equal $N(\pi \cdot \delta \cdot \Pi^m)$.

[22] I assume the upstream technologies (genes) do not compete in the same product market as the downstream technologies (diagnostic tests).

Social welfare in this case will be the product of three terms: first, the fraction π of firms who enter that will successfully discover an innovation; second, the total amount of entry $N(\pi \cdot \delta \cdot \Pi^m)$; and third, a weighted sum of the level of social welfare generated under monopoly pricing (weighted by δ) and the level of social welfare generated under marginal cost pricing (weighted by $1 - \delta$). Denoting the choke-off price (that is, the lowest price at which there is no demand) by $\bar{p}$, social welfare under monopoly pricing is equal to the following sum of consumer surplus and producer surplus:

$$SW^m = \int_{p^m}^{\bar{p}} D(p)dp + \Pi^m \quad (6)$$

Social welfare under marginal cost pricing is:

$$SW^c = \int_0^{\bar{p}} D(p)dp \quad (7)$$

We can thus write social welfare in the case in which the upstream technologies are in the public domain as:

$$SW^{public} = \pi \cdot N(\pi \cdot \delta \cdot \Pi^m) \cdot [\delta \cdot SW^m + (1-\delta) \cdot SW^c] \quad (8)$$

Next, consider the second case in which the upstream technologies are granted patents. I assume all downstream products are "cumulative" in the sense of infringing on the upstream firm's patents, so that any firms developing downstream products must obtain a license from the upstream firm. I assume that patents on the upstream technology ensure there will be no generic entry in the downstream market (since generics would also infringe on the upstream firm's patents). However, because licensing negotiations are assumed to occur *ex post* (that is, after the downstream product developer has already sunk her R&D costs), there is a hold-up problem where the upstream firm can extract rents after a downstream product is discovered. I capture the reduced form outcome of these *ex post* negotiations by assuming the downstream firm captures a share λ of the profits from the downstream product. Firms will enter as long as their cost c is less than the share of expected monopoly profits they would capture, $\pi \cdot \lambda \cdot \Pi^m$, so the total amount of entry will equal $N(\pi \cdot \lambda \cdot \Pi^m)$.

Social welfare in this case will again be the product of three terms: first, the fraction π of firms who enter that will successfully discover an innovation; second, the total amount of entry $N(\pi \cdot \lambda \cdot \Pi^m)$; and third, the level of social welfare generated under monopoly pricing. We can thus write social welfare in the case in which the upstream technologies are granted patents as:

$$SW^{ip} = \pi \cdot N(\pi \cdot \lambda \cdot \Pi^m) \cdot SW^m \quad (9)$$

Comparing the amount of entry in the two cases is straightforward: there is more entry when the upstream technologies are in the public domain, relative to when they are granted

patents, if and only if $\delta > \lambda$.[23] One can also straightforwardly derive expressions for the level of social welfare in both cases: there is a cutoff value of λ, denoted λ^*, such that social welfare from the upstream technologies being in the public domain is the same as social welfare from the upstream technologies being granted patents. For values $\lambda < \lambda^*$, social welfare is strictly higher when the upstream technologies are in the public domain relative to when they are granted patents. For values $\lambda > \lambda^*$, social welfare is strictly higher when the upstream technologies are granted patents relative to when they are in the public domain. The value of λ^* is greater than δ.

5.2 Empirics

5.2.1 Survey evidence—A series of surveys by Cohen, Walsh, and colleagues have documented evidence on how patents affect follow-on research investments. Walsh, Arora and Cohen (2003) interviewed 70 individuals in various positions related to research and intellectual property in biomedical and pharmaceutical research.[24] The goal of the survey was to elicit responses on whether increasingly broad patents or increasing complexity in navigating patents on research tools have impeded follow-on research. Survey respondents confirmed (uniformly) that the patent landscape had indeed become more complex in recent years, with more patents per innovation (including patents on research tools) and increased patenting of upstream technologies (such as targets for drug discovery). However, almost no respondents reported abandoning worthwhile projects because of issues of access to intellectual property. Rather, both university and industrial researchers reported adopting various “working solutions” that allowed their research to proceed, including licensing, inventing around patents, going offshore, the development and use of public databases and research tools, court challenges, and simply using the technology without a license (i.e. infringement). Industrial intellectual property holders who addressed the issue reported tolerating academic research infringing on their intellectual property on research tools (with the exception of patents on diagnostic tests used in clinical research), partly because it can increase the value of the patented technology. Industrial respondents also noted that the small prospective gains from lawsuits would not be worth the legal fees, the risk of the patent being narrowed or invalidated, and the bad publicity from suing a university.

Walsh, Cohen and Cho (2007) conducted a similar survey of academic biomedical researchers, investigating the impact of patents on access to the knowledge and material inputs used in follow-on research.[25] Table 3 documents tabulations of these survey responses, showing the reasons why academics say they choose not to pursue research projects. In general, terms demanded for access to needed research inputs (10%; “unreasonable terms”) and too many patents (3%) are relatively rarely cited as reasons for project abandonment. Unreasonable terms are somewhat more frequently cited as a determinant of research project choice for research related to drug discovery (21%) relative to basic research (9%). Other survey results (not reported) suggest that only 5% of survey

[23]See appendix for a formal derivation of this result.
[24]Interviewees included intellectual property attorneys, scientists and managers at ten pharmaceutical and fifteen biotechnology firms, academic researchers and technology transfer officers at six universities, and other intellectual property attorneys and personnel working in government and trade associations.
[25]The authors conducted a survey of 1125 biomedical researchers in universities, government labs, and non-profits; they received 414 responses. They also surveyed a second sample of 270 researchers working on signal proteins, from which they received 93 responses.

respondents report regularly checking for patents on knowledge or tangible inputs related to their research, suggesting that – consistent with the findings of Walsh, Arora and Cohen (2003) – many of the surveyed academics may simply be ignoring patents.

5.2.2 Evidence from judges and patent examiners—Moving beyond survey evidence, two empirical challenges must be overcome in order to estimate how patents on existing technologies affect follow-on innovation. First, because we expect that the process through which some technologies are patented and others are not is non-random, identifying a causal effect of patents on follow-on innovation requires isolating a quasi-experiment in which patent protection is as good as randomly assigned to some technologies and not others. Second, even though there is widespread agreement that follow-on innovation is quite common – in the sense that many technologies build on existing technologies – it is challenging to construct data that can trace these links between technologies and follow-on innovations in practice. Two recent papers, Galasso and Schankerman (2015) and Sampat and Williams (2015), attempt to circumvent each of these empirical challenges.

Galasso and Schankerman (2015) focus on the empirical context of patent validity cases reviewed by the U.S. Court of Appeals for the Federal Circuit. This setting is empirically useful because they argue that judges are assigned to patent cases through a computer program that randomly generates three-judge panels, with decisions governed by majority rule. This random assignment, together with the fact that judges vary in their propensity to invalidate patents, allows for a novel instrumental variables approach in which the decision of the court on a given case can be instrumented with the predicted propensity to invalidate of the assigned three-judge panel. Because the judge panels are randomly assigned, this instrument should only affect follow-on innovation through affecting the validity ruling on the case.

Among patents that reach the Federal Circuit court, some are upheld and others are invalidated. In order to measure the (instrumented) effect of patent invalidation on follow-on innovation, Galasso and Schankerman measure subsequent citations to the litigated patent. They motivate this measure by noting that because patents constitute prior art, subsequent innovators are still required to cite patents when relevant even if those patents have been declared invalid and the underlying technologies have moved into the public domain.[26] Their baseline estimates suggest that (instrumented) patent invalidation leads to about a 50 percent increase in subsequent citations to the focal patent. Figure 5 documents the IV estimates of the effect of patent invalidation in each of the 10 years following the Federal Circuit decision, along with the associated 90% confidence intervals. This graph suggests that the effects of invalidation on subsequent citations persists for seven years after the decision.

By construction, this empirical strategy can only be applied to patents challenged in cases that reach the Federal Circuit court. While certainly a selected value of patents, this sample covers patents in a relatively diverse set of technology fields, allowing Galasso and

[26]For two of the technology fields in their sample, the authors also construct alternative non-citation based measures of follow-on innovation. The findings they document for these alternative outcome measures are consistent with those uncovered using their citation-based outcomes, so I do not here focus on discussing these measures.

Schankerman to explore the heterogeneity of their results across different types of technologies. These results suggest that patent invalidation has an impact on follow-on innovation in some fields such as computers and communications, electronics, and medical instruments, but not in other fields such as drugs, chemicals, and mechanical technologies. They also present results suggesting that the impact of patent rights on follow-on innovation is driven completely by the invalidation of patents owned by large firms, and by increases in the number of small innovators subsequently citing the focal patent.

Sampat and Williams (2015) focus on a case study of one particular technology – patents on human genes. They start with the sample of published USPTO patent applications claiming intellectual property over human genes. In recent years, USPTO patent applications claiming human genes were required to list the DNA sequence of the relevant gene(s) in the text of the patent. By applying bioinformatics methods as in Jensen and Murray (2005) it is possible to link the DNA sequences claimed in patents with a standard set of gene identifiers, which can in turn be linked to measures of follow-on innovation on those genes. Specifically, Sampat and I construct data measuring the scientific publications related to each gene (an indicator of scientific research investments), and measure the use of genes in pharmaceutical clinical trials and medical diagnostic tests (two indicators of commercial research investments). Because gene patents have been widely perceived as sufficiently broad that these follow-on innovations would require gene patent licenses (see, e.g., Heller and Eisenberg (1998)), this data construction corresponds closely with the type of follow-on innovation described theoretically in the literature.

Using this data, we can compare the amount of follow-on innovation on human genes that are patented and not patented. However, as in Galasso and Schankerman, we are concerned that gene patents are likely not randomly assigned across genes. Indeed, a naive comparison of follow-on innovation across patented and non-patented genes in our data suggests that we observe higher levels of follow-on innovation on patented genes, but that is true both before and after these genes are patented – highlighting a selection bias in which genes are patented.

To address this selection bias concern, we construct two quasi-experimental approaches. First, we present a simple comparison of follow-on innovation across genes claimed in accepted and “rejected” patent applications. As described in Section 2, patents are not actually rejected by the USPTO, but we can compare genes claimed in patent applications that were granted patents with genes claimed in patent applications but never granted a patent – which is useful if genes that were included in unsuccessful patent applications can serve as a valid comparison group for genes that were granted patents. Second, we develop a novel instrumental variables approach in which the existence of a patent on a given gene is instrumented with the “leniency” of the assigned patent examiner.[27] Previous qualitative research has suggested that the assignment of patent applications to patent examiners at the USPTO is plausibly random conditional on some covariates such as technology type and application year (Cockburn, Kortum and Stern, 2003; Lemley and Sampat, 2010), and we

[27]Note that this instrumental variables strategy has subsequently been re-applied in a number of other papers, including Gaule (2015) and Farre-Mensa, Ljungqvist and Hegde (2016).

confirm empirically that this assumption looks plausible in our sample of gene patent applications – suggesting that examiner leniency can serve as a valid instrument.

Both of our quasi-experimental approaches suggest that gene patents have not had quantitatively important effects on follow-on innovation. Figure 6 graphically documents the main results from the first quasi-experimental approach. Among human genes included in at least one patent application, about 30 percent are included in a granted patent. Figure 6(a) documents the time path of when genes in the patented group receive their (first) patent grant over time: about half have received a patent grant by 2005, giving some sense of when we might expect any differences in follow-on innovation between the two groups to be observed. Figures 6(b) and 6(c) document trends in follow-on innovation by year. Comfortingly, the number of scientific publications (Figure 6(b)) and the number of clinical trials (Figure 6(c)) look quite similar across genes claimed in accepted and rejected patent applications prior to 2001, the year when our data on gene patent applications starts.[28] But the two data series also look quite similar post-2001, providing little evidence that patent grants have had any effect on either follow-on innovation outcome.

The estimates from our first approach – comparing genes claimed in accepted and rejected patent applications – are reasonably precise, and are able to reject the types of declines in follow-on innovation that have been observed in quasi-experimental studies of non-patent forms of intellectual property (Murray, Aghion, Dewatripont, Kolev and Stern, 2008; Williams, 2013). While the estimates from the second instrumental variables approach are less precise, the fact that both approaches generate similar conclusions is comforting. Most directly, these estimates speak against the arguments made by the US Supreme Court in the case *Association for Molecular Pathology v. Myriad Genetics*, in which the Court unanimously ruled to invalidate patents on human genes, arguing that such patents *"would 'tie up'…[genes] and…inhibit future innovation."* More broadly, our analysis is complementary to that of Galasso and Schankerman in providing evidence on when and how patents will affect follow-on innovation.[29]

6 Conclusion

The patent system is a widely-used policy lever attempting to better align the private returns to developing new technologies with the social value of those inventions. The past few decades have seen the development of large academic literatures in a variety of fields – including economics, law, and strategy, among others – investigating various aspects of the patent system. However, surprisingly little research has focused on empirically estimating the key parameters needed to evaluate the social costs and social benefits of the patent system. A half-century ago, Penrose (1951) and Machlup (1958) argued that insufficient empirical evidence existed to make a conclusive case either for or against patents. Today, I would argue that given the limitations of the existing literature we still have essentially no

[28] In contrast, the level of follow-on innovation on genes never included in a patent application are notably lower throughout the sample. While these genes are not part of our empirical analysis, that pattern is consistent with strong selection of genes into patent *filings*.

[29] Consistent with our results, Galasso and Schankerman find no evidence that patent invalidations increase patent citations in the NBER technology field that includes most of our gene patents ("drugs and medical").

credible empirical evidence on the seemingly simple question of whether stronger patent rights – either longer patent terms or broader patent rights – encourage research investments into developing new technologies. While researchers have recently begun to make progress on the more limited question of how patents on existing technologies affect follow-on innovation (Galasso and Schankerman, 2015; Sampat and Williams, 2015), evidence on the overall effects of patents on research investments are needed as one input into optimal patent policy design.

Some of the evidence discussed in this article suggests there may be important heterogeneity across industries or across technologies in how patents affect research investments. While legal constraints limit the ability to which current patent law could be tailored away from the current "one size fits all" reward approach, researchers such as Roin (2014) have argued that such tailoring may be feasible at least in principle given observable economic determinants of optimal patent strength. Development of additional empirical evidence informing the question of optimal patent terms and how they vary across industries is an important direction for future work, which could quantify the potential benefits of such tailored patent awards.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Arrow, Kenneth. Economic welfare and the allocation of resources for invention. In: Nelson, Richard, editor. The Rate and Direction of Inventive Activity. Princeton University Press; 1962.

Bessen, James. Holdup and licensing of cumulative innovations with private information. Economics Letters. 2004; 82(3):321–326.

Bessen, James, Maskin, Eric. Sequential innovation, patents, and imitation. RAND Journal of Economics. 2009; 40(4):611–635.

Bloom, Nicholas, Schankerman, Mark, Van Reenen, John. Identifying technology spillovers and product market rivalry. Econometrica. 2013; 81(4):1347–1393.

Boldrin, Michele, Levine, David K. The case against patents. Journal of Economic Perspectives. 2013; 27(1):3–22.

Budish, Eric, Roin, Benjamin, Williams, Heidi. Do firms underinvest in long-term research? Evidence from cancer clinical trials. American Economic Review. 2015; 105(7):2044–2085. [PubMed: 26345455]

Budish, Eric, Roin, Benjamin, Williams, Heidi. Patents and research investments: Assessing the empirical evidence. American Economic Review Papers & Proceedings. 2016; 106(5):183–187.

Carley, Michael, Hegde, Deepak, Marco, Alan. What is the probability of receiving a US patent? Yale Journal of Law & Technology. forthcoming.

Cockburn, Iain M., Lanjouw, Jean O., Schankerman, Mark. Patents and the Global Diffusion of New Drugs. American Economic Review. Jan; 2016 106(1):136–64.

Cockburn, Iain, Kortum, Samuel, Stern, Scott. Are all patent examiners equal? Examiners, patent characteristics, and litigation outcomes. In: Cohen, Wesley, Merrill, Stephen, editors. Patents in the Knowledge-Based Economy. National Academies Press; 2003.

Cohen, Wesley M., Goto, Akira, Nagata, Akiya, Nelson, Richard R., Walsh, John P. R&D spillovers, patents and the incentives to innovate in Japan and the United States. Research policy. 2002; 31(8): 1349–1367.

Cohen, Wesley M., Nelson, Richard R., Walsh, John P. Technical Report. National Bureau of Economic Research; 2000. Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not).

Cutler, David. Your Money or Your Life: Strong Medicine for America's Health Care System. Oxford University Press; 2004.

Duggan, Mark, Garthwaite, Craig, Goyal, Aparajita. The Market Impacts of Pharmaceutical Product Patents in Developing Countries: Evidence from India. American Economic Review. Jan; 2016 106(1):99–135.

Farre-Mensa, Joan, Ljungqvist, Alexander, Hegde, Deepak. The Bright Side of Patents. NYU-Stern. 2016 mimeo.

Galasso, Alberto, Schankerman, Mark. Patents and cumulative innovation: Causal evidence from the courts. Quarterly Journal of Economics. 2015; 130(1):317–369.

Gans, Joshua S., Hsu, David H., Stern, Scott. The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. Management Science. 2008; 54(5):982–997.

Gaule, Patrick. Patents and the Success of Venture-Capital Backed Startups: Using Examiner Assignment to Estimate Causal Effects. CERGE-EI. 2015 mimeo.

Gilbert, Richard, Shapiro, Carl. Optimal patent length and breadth. RAND Journal of Economics. 1990; 21:106–112.

Graham, Stuart, Hegde, Deepak. Disclosing patents' secrets. Science. 2015; 347(6219):236–237. [PubMed: 25593175]

Green, Jerry, Scotchmer, Suzanne. On the division of profit in sequential innovation. RAND Journal of Economics. 1995; 26(1):20–33.

Haber, Stephen. Patents and the Wealth of Nations. George Mason Law Review. 2016; 23(4)

Hall, Bronwyn H., Jaffe, Adam B., Trajtenberg, Manuel. Technical Report. National Bureau of Economic Research; 2001. The NBER patent citation data file: Lessons, insights and methodological tools.

Hegde, Deepak, Luo, Hong. Patent Publication and the Market for Ideas. Management Science. forthcoming.

Heller, Michael, Eisenberg, Rebecca. Can patents deter innovation? The anticommons in biomedical research. Science. 1998; 280(5364):698–701. [PubMed: 9563938]

Jensen, Kyle, Murray, Fiona. Intellectual property landscape of the human genome. Science. 2005; 310(5746):239–240. [PubMed: 16224006]

Kitch, Edmund. The nature and function of the patent system. Journal of Law and Economics. 1977; 20:265–290.

Klemperer, Paul. How broad should the scope of patent protection be? RAND Journal of Economics. 1990; 21:113–130.

Kortum, Samuel, Lerner, Josh. What is behind the recent surge in patenting? Research policy. 1999; 28(1):1–22.

Kyle, Margaret, Qian, Yi. National Bureau of Economic Research (NBER) working paper #20799. 2014. Intellectual property rights and access to innovation: Evidence from TRIPS.

Lemley, Mark, Sampat, Bhaven. Is the patent office a rubber stamp? Emory Law Journal. 2008; 58:181–203.

Lemley, Mark, Sampat, Bhaven. Examining patent examination. Stanford Technology Law Review. 2010; 2

Lemley, Mark, Shapiro, Carl. A Simple Approach to Setting Reasonable Royalties for Standard-Essential Patents. Berkeley Technology Law Journal. 2013; 28:1135–1166.

Lerner, Josh. The empirical impact of intellectual property rights on innovation: Puzzles and clues. American Economic Review Papers & Proceedings. 2009; 99(2):343–348.

Levin, Richard C., Klevorick, Alvin K., Nelson, Richard R., Winter, Sidney G. Appropriating the returns from industrial research and development. Brookings Papers on Economic Activity. 1987; 1987(3):783–831.

Machlup, Fritz. An economic review of the patent system number 15. US Government Printing Office; 1958.

Mansfield, Edwin. Patents and innovation: An empirical study. Management Science. 1986; 32(2): 173–181.

Marco, Alan C., Sarnoff, Joshua D., DeGrazia, Charles A. Technical Report. Hoover Institution Working Group on Intellectual Property, Innovation, and Prosperity; 2016. Patent claims and patent scope.

Murray, Fiona, Aghion, Philippe, Dewatripont, Mathias, Kolev, Julian, Stern, Scott. Of mice and academics: Examining the effect of openness on innovation. 2008 unpublished MIT mimeo.

Nelson, Richard. The simple economics of basic scientific research. Journal of Political Economy. 1959; 67(3):297–306.

Nordhaus, William. Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change. MIT Press; 1969.

Nordhaus, William. The health of nations: The contribution of improved health to living standards. In: Murphy, Kevin M., Topel, Robert H., editors. Measuring the Gains from Medical Research: An Economic Approach. Vol. chapter 1. Chicago, IL: University of Chicago Press; 2003. p. 9-40.

Ouellette, Lisa Larrimore. Do patents disclose useful information? Harvard Journal of Law & Technology. 2012; 532(2):531–593.

Pakes, Ariel. Patents as options: Some estimates of the value of holding European patent stocks. Econometrica. 1986; 54(4):755–784.

Penrose, Edith Tilton. The Economics of the International Patent System number 30. Baltimore, Md: Johns Hopkins Press; 1951.

Plant, Arnold. The economic theory concerning patents for inventions. Economica. 1934; 1(1):30–51.

Putnam, Jonathan. Yale PhD dissertation. 1996. The value of international patent protection.

Roin, Benjamin. The disclosure function of the patent system (or lack thereof). Harvard Law Review. 2005; 118(6):2007–2028.

Roin, Benjamin. The case for tailoring patent awards based on time-to-market. UCLA Law Review. 2014; 61(672)

Sakakibara, Mariko, Branstetter, Lee. Do stronger patents induce more innovation? Evidence from the 1998 Japanese patent law reforms. RAND Journal of Economics. 2001; 32(1):77–100.

Sampat, Bhaven, Williams, Heidi. How do patents affect follow-on innovation? Evidence from the human genome. MIT mimeo. 2015

Schankerman, Mark, Pakes, Ariel. Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period. The Economic Journal. 1986; 96(384):1052–1076.

Schmookler, Jacob. Invention and Economic Growth. Cambridge, MA: Harvard University Press; 1966.

Scotchmer, Suzanne. Standing on the shoulders of giants: Cumulative research and the patent law. Journal of Economic Perspectives. 1991; 5(1):29–41.

Scotchmer, Suzanne. Innovation and Incentives. MIT Press; 2004.

Walsh, John, Arora, Ashish, Cohen, Wesley. Working through the patent problem. Science. 2003; 299(5609):1021. [PubMed: 12586928]

Walsh, John, Cohen, Wesley, Cho, Charlene. Where excludability matters: Material versus intellectual property in academic biomedical research. Research Policy. 2007; 36(8):1184–1203.

Williams, Heidi. Intellectual property rights and innovation: Evidence from the human genome. Journal of Political Economy. 2013; 121(1):1–27.

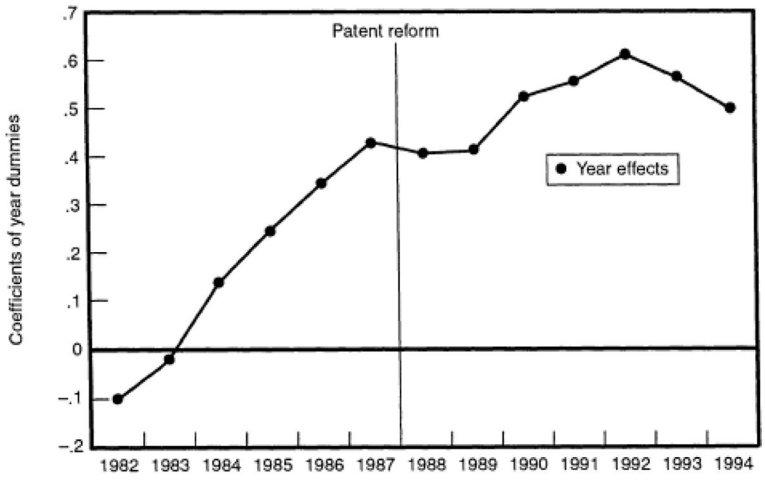

Figure 1.

Patent law changes and research investments: Evidence from Sakakibara and Branstetter (2001)

Source: Sakakibara and Branstetter (2001) Figure 5.

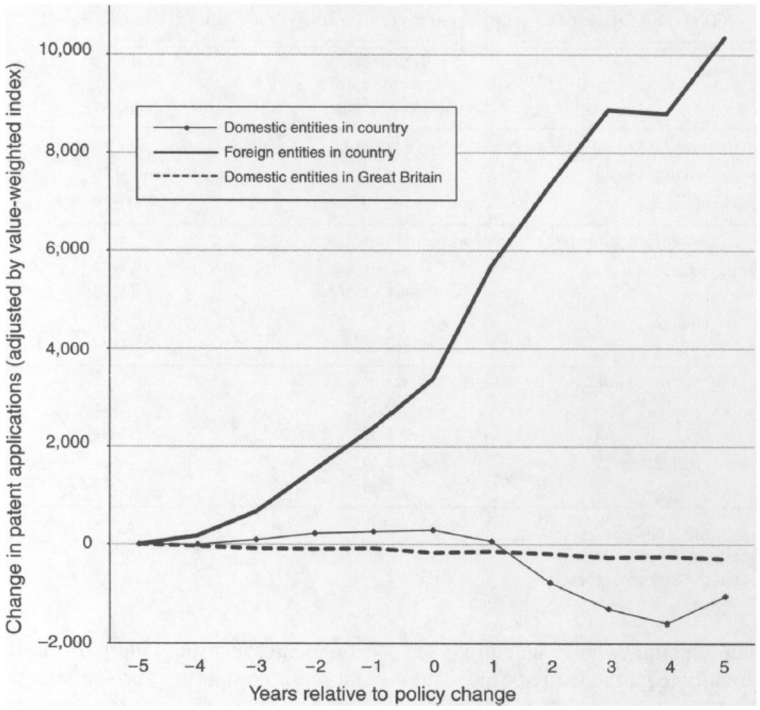

Figure 2.

Patent law changes and research investments: Evidence from Lerner (2009)

Source: Lerner (2009) Figure 2.

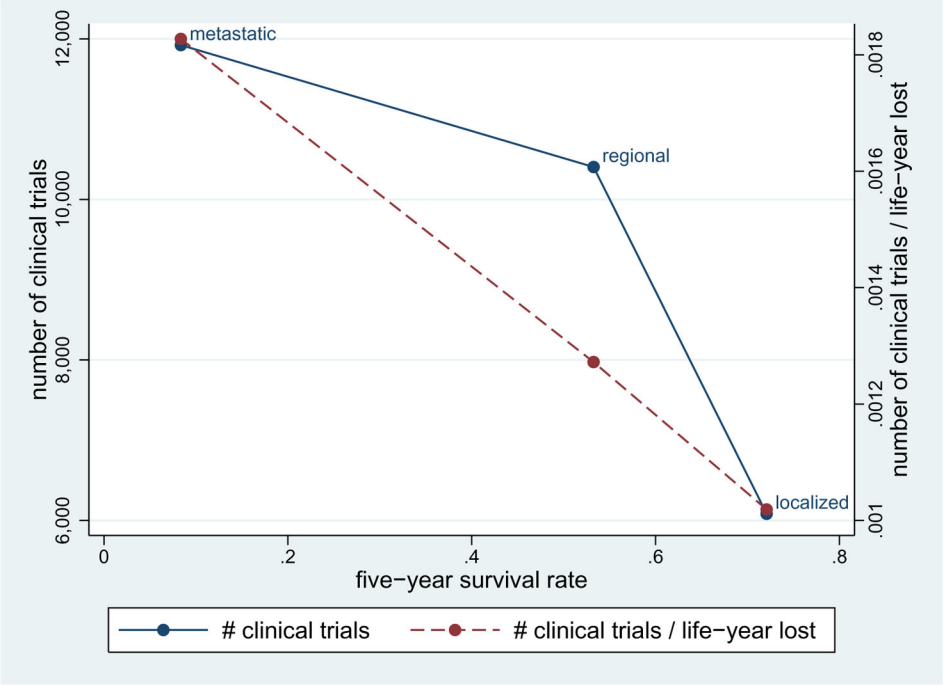

Figure 3.

Research investments and clinical trial length: Evidence from Budish, Roin and Williams (2015)

Source: Budish, Roin and Williams (2015) Figure 1 Panel A.

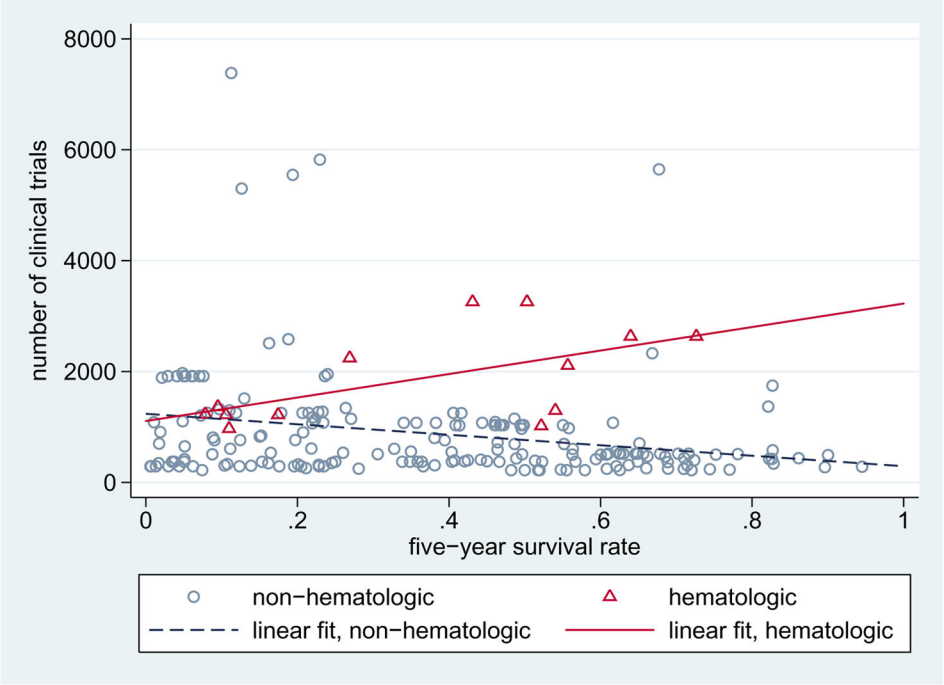

Figure 4.

Research investments and surrogate endpoints: Evidence from Budish, Roin and Williams (2015)

Source: Budish, Roin and Williams (2015) Figure 4.

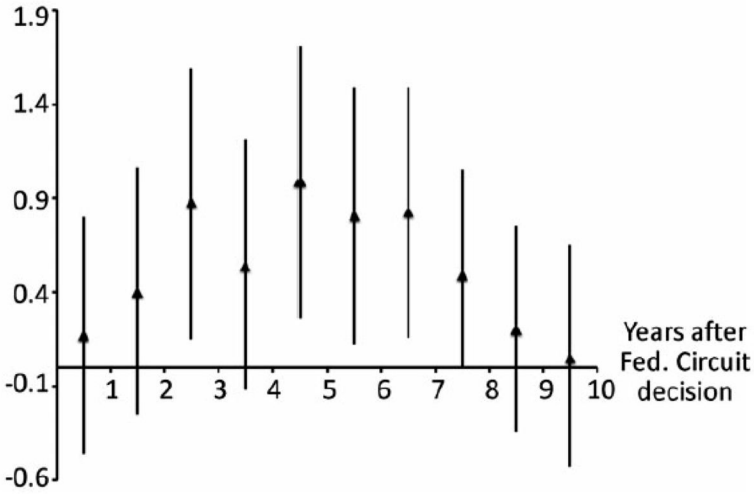

Figure 5.

Patents and cumulative innovation: Evidence from Galasso and Schankerman (2015)

Source: Galasso and Schankerman (2015) Figure II.

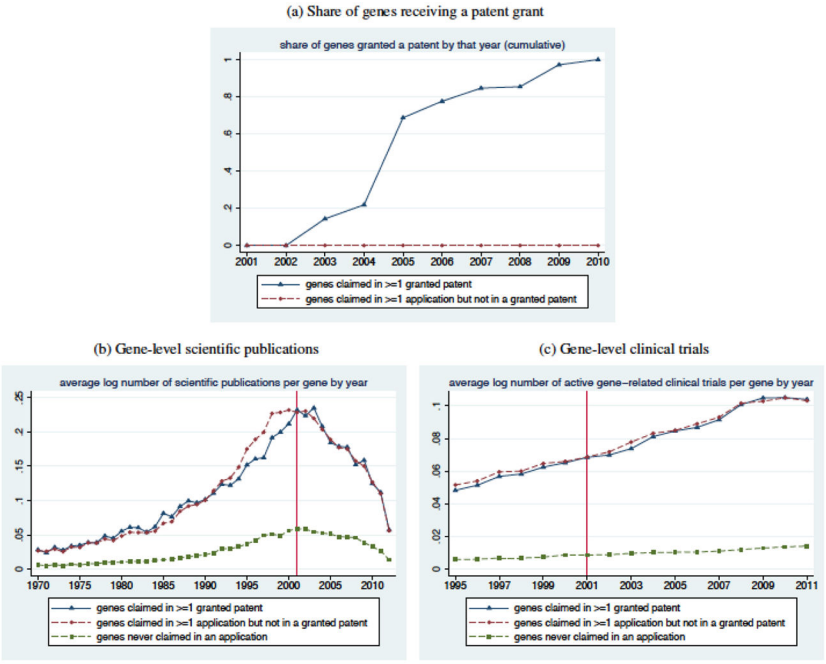

Figure 6.

Patents and cumulative innovation: Evidence from Sampat and Williams (2015)

Source: Sampat and Williams (2015) Figure 2.

Table 1

Patents and research investments: Evidence from Mansfield (1986) survey

***Percent of Developed or Commercially Introduced Inventions That Would Not Have Been Developed or Commercially Introduced if Patent Protection Could Not Have Been Obtained, Twelve Industries*, 1981–83.[a]**

Industry	Percent That Would Not Have Been Introduced	Percent That Would Not Have Been Developed
Pharmaceuticals	65	60
Chemicals	30	38
Petroleum	18	25
Machinery	15	17
Fabricated metal products	12	12
Primary metals	8	1
Electrical equipment	4	11
Instruments	1	1
Office equipment	0	0
Motor vehicles	0	0
Rubber	0	0
Textiles	0	0

Source: Mansfield (1986) Table 1.

Table 2

Patents and appropriability: Evidence from Yale survey

	Process patents		Product patents	
Industry	**Mean**	**Standard error**	**Mean**	**Standard error**
Pulp, paper, and paperboard	2.6	0.3	3.3	0.4
Cosmetics	2.9	0.3	4.1	0.4
Inorganic chemicals	4.6	0.4	5.2	0.3
Organic chemicals	4.1	0.3	6.1	0.2
Drugs	4.9	0.3	6.5	0.1
Plastic materials	4.6	0.3	5.4	0.3
Plastic products	3.2	0.3	4.9	0.3
Petroleum refining	4.9	0.4	4.3	0.4
Steel mill products	3.5	0.7	5.1	0.6
Pumps and pumping equipment	3.2	0.4	4.4	0.5
Motors, generators, and controls	2.7	0.3	3.5	0.5
Computers	3.3	0.4	3.4	0.4
Communications equipment	3.1	0.3	3.6	0.3
Semiconductors	3.2	0.4	4.5	0.4
Motor vehicle parts	3.7	0.4	4.5	0.4
Aircraft and parts	3.1	0.5	3.8	0.4
Measuring devices	3.6	0.3	3.9	0.3
Medical instruments	3.2	0.4	4.7	0.4
Full sample	3.5	0.06	4.3	0.07

Source: Levin, Klevorick, Nelson and Winter (1987) Table 2.

Table 3

Patents and follow-on research: Survey evidence from Walsh, Cohen and Cho (2007)

Reasons for not pursuing projects, by research goal and for signal proteins samples								
		Random sample	Research goal			Signal proteins samples		
			Drug discovery	Basic research	Other	CTLA-4	EGF	NF-kB
No funding	%High	62	86	60	58	63	54	82
Too busy	%High	60	55	60	59	53	58	48
Not feasible	%High	46	41	46	47	33	55	53
Not scientifically important	%High	40	24	41	45	40	36	50
Not interesting	%High	35	24	36	33	20	30	29
Too much competition	%High	29	21	32	21	27	29	29
Little social benefit	%High	15	21	14	15	13	5	22
Unreasonable terms	%High	10	21	9	6	7	9	19
Not help w/promotion/job	%High	10	21	7	15	0	13	5
Too many patents	%High	3	3	2	3	0	4	0
New firm unlikely	%High	3	3	2	3	0	4	0
Little commercial potential	%High	2	3	2	3	0	4	0
Little income potential	%High	1	3	1	3	0	4	0
Not patentable	%High	1	3	1	3	0	4	0
Respondents	*N*	274	28	213	33	16	24	22

Note: "%High" is the percent answering "4" or "5" on a five point scale ranging from "1: not at all important" to "5: very important".

Source: Walsh, Cohen and Cho (2007) Table 3.