

Patent Value and R&D Competition

Laura Magazzini*, Fabio Pammolli†
Massimo Riccaboni‡§

PRELIMINARY DRAFT

Please do not quote without author permission

1 Introduction

The pharmaceutical industry faces continued criticism over the productivity of R&D spend and in particular over the value of “me too” innovation. There is a lack of understanding of the nature of scientific advance in medicine and the extent of uncertainty (in terms of project failure) of R&D work in a therapy area.

This paper – which is part of a larger on-going research project on the properties of R&D competition in pharmaceuticals – looks at the features of the learning process that characterizes the actors operating within the pharmaceutical domain. This industry represents a unique framework for studying issues related to innovation and innovative activities, given its strong roots into the realm of scientific knowledge, the important role played by patents as a means for protecting economic returns from R&D (in exchange of the full disclosure of the characteristics of the innovation), the high level of competition and its distinctive industry structure, where actors with different ethos, especially with regard to information sharing and disclosure,

*University of Verona

†University of Florence and IMT, Lucca

‡University of Florence

§This paper is the result of the “R&D competition” research project carried out jointly with Adrian Towse and Martina Garau of the Office of Health Economics, London, UK. A preliminary draft of the paper has been presented to the DRUID Summer Conference 2006 (Copenhagen), to the 11th ISS Conference (Sophia-Antipolis), and to the IHEA conference, Sixth World Congress. Comments and suggestions by Fabio Montobbio, Mark Supekar, Cristina Tinti are gratefully and conference discussants and participants are gratefully acknowledged.

and different capabilities coexist and have created a wide network of collaborations¹.

As a result of the strong linkages that exist between drug development and the scientific advances in the “open science”, firms dissect and analyze an increasing number of techniques, trajectories and exploration strategies (Orsenigo, Pammolli and Riccaboni 2001). Even if pooling from a common ground, research investments across firms are weakly correlated, after removing the common effect due to exogenous shocks (Henderson and Cockburn 1994). This pattern notwithstanding, knowledge spillovers play a significant role in pharmaceutical research and competing projects exhibit complementary patterns, as rival research results are positively correlated with firm productivity (Henderson and Cockburn 1994, Henderson and Cockburn 1996).

Against this background, we build on a comprehensive dataset about the innovative activity of pharmaceutical and biotechnology firms, including R&D project level data, patents, citations, and collaborations, and explore the nature of technological advances and of the underlying technological (learning) regime, shaping the industrial patterns of innovative activity (Nelson and Winter 1982, Winter 1994, Malerba and Orsenigo 1993, Breschi, Malerba and Orsenigo 2000).

The literature has analyzed the pattern of successes over time, finding evidence of low serial correlation in the introduction of successful products within families of chemically related compounds at the firm level (Sutton 1998). We take a broader perspective here and consider research efforts spanning both from R&D failures and successes.

Given the high relevance of patent for protecting pharmaceutical innovations (Cohen, Nelson and Walsh 2000, Arundel and Kabla 1998), research efforts are proxied using patents, whereas knowledge utilization and spillovers are measured by looking at the pattern of patent citations². The key assumption is that a citation made to a previous patent denotes a knowledge transfer from the cited patent to the citing one³. Patents rule out direct imitation of the innovative compound or process, nonetheless the information disclosed through patents expand the knowledge frontier and can provide

¹See, e.g., Powell, Koput and Smith-Doerr (1996)

²See Jaffe and Trajtenberg (2002) and the literature referenced therein.

³We are aware that patent citation count is only a noisy proxy of the relevance of the knowledge disclosed, since citations might be included for strategic purposes or added by firm’s lawyers or by patent examiners (Alcacer and Gittelman 2004). However, survey evidence shows that, even if noisy, patent citations are indicative of knowledge spillovers and communication among inventors (Jaffe, Trajtenberg and Fogarty 2000).

rivals useful insights into new chemical and pharmacological properties of compounds or mechanism of action, eventually leading to new patentable compounds or processes.

Our analysis reveals that technological competencies are accumulated not only building on successfully developed compounds, that nonetheless play a substantial role in guiding subsequent research efforts of both the innovating firm and its rivals, but also on failures, i.e. compounds that do not pass through all the stages involved in drug development, rather they are discontinued due, for example, to lack of effectiveness or toxicological effects. Actually patents whose knowledge base also comprises failures by rival firms have higher relevance as measured by the number of life-time received citations.

The distinction between leadlike and druglike compounds is useful in interpreting our results. At the early stages of drug development, firms identify lead structures, i.e. compounds that typically exhibit sub-optimal target binding affinity, but with relatively simple chemical features, well-established structure-activity relationship, good properties in terms of absorption, distribution, metabolism, and excretion, and a favorable patent situation, making them good starting points in medicinal chemistry efforts (Oprea, Davis, Teague and Leeson 2001). Even though, they will never reach the market, a large number of subsequent development builds on their structure.

In the pharmaceutical technological paradigm (Dosi 1988), firms need to be able to master knowledge from many different sources when searching for new molecules with optimal target binding affinity properties. First basic scientific knowledge about the relationship between chemical structures and physical properties, then technological capabilities and previous experience, built on the basis of both its failures and successes, and of failures and successes of rival firms. The analysis presented in the paper shows that knowledge about failed compounds play an important role in guiding subsequent research efforts, both within and outside the originating firm boundaries.

The paper is organized as follows. Section 2 describes the features of the drug development process, which is characterized by high uncertainty, low cumulateness, and by a large presence of knowledge spillovers. Section 3 describes the data and the methods used in this study. Section 4 presents the empirical results, discussed in Section 5, also drawing implications in terms of efficiency of the research efforts at the sectoral level.

2 The Drug Development Process

When talking about the innovation process and its dynamic properties, the pharmaceutical industry is a peculiar one in many respects.

The pharmaceutical industry is a textbook example of a “science-based” sector (Pavitt 1984), where innovation, both in the form of new therapeutic products and improvements of existing products (in terms of better delivery, reduced side effects, or improved efficacy) is jointly driven by advances in the field of applied sciences and in the knowledge about bacterial, animal and human processes led by the scientific community. Innovation is, in turn, the fundamental source of firm competitiveness and profitability.

However, the R&D process in pharmaceuticals involve high costs and is subject to high uncertainty. On the basis of the internal records of ten large pharmaceutical companies, an average 802 million US dollar pre-approval cost estimate is reported (DiMasi, Hansen and Grabowski 2003). In addition, a significant share of R&D projects is abandoned due to the emergence of toxicological effects or to the lack of effectiveness in treating the targeted disease. FDA estimates that only a small percentage of the discovered compounds lead to a marketable product: among the compounds selected for human clinical trials⁴, 70% passes Phase I, while the share of successful compounds is significantly reduced in the case of Phase II and Phase III, respectively 33% and 25-30% (Trenter 1999). Even if a product successfully reaches the market, firms do not suspend the monitoring activity to check the emergence of side effects or new toxicological evidence that might eventually lead to the withdrawal of the product from the market⁵.

In addition, the pharmaceutical R&D process is characterized by a large presence of R&D spillovers. Larger firms enjoy higher productivity rates not only for economies of scale, but also for economies of scope spanning from a high level of diversification in the R&D activity that allow firms to capture

⁴The innovative process of drug development evolves linearly through well-defined stages. When a firm discovers a potentially active substance a patent is applied for preventing firm’s rivals from copying the new compound or technology. Even if important, the patent is only the first step in a lengthy process for the development of new drugs, not always ending with a product that can be commercialized on the market. First, preclinical trials are aimed at assessing the safety of administering the compound to human, and then clinical trials are carried over for assessing its safety and effectiveness in targeting the selected indications. Only in case all the stages are successfully passed, the new compound can be registered and then marketed, allowing the firm to recover the R&D costs necessary for its development.

⁵However, monitoring side effects (emerging also during clinical trials) might also produce evidence suggesting new application of a compound, the most striking example being Viagra (Kling 1998).

internal and external knowledge spillovers. Projects in the same field or in the same therapeutic category undertaken by competitors exert a positive impact on the firms' research productivity, as measured by the number of important patents obtained from a given research program (Henderson and Cockburn 1996).

Also high firm turnover characterizes both the R&D and the market sides. In an analysis of the patents related to the memory-enhancing agents (MEA), a "tree-plane" model of technological development is presented. The authors identify the central technology plane, containing patents related to MEA, and its precursor and successor technology planes, containing respectively the earlier, cited research, and the patent protecting the new applications or variations of the MEA central technology. The three planes are populated by different sets of institutions, showing that firms other than the central MEA innovator are able to catch the new technological opportunities (Narin, Smith jr. and Albert 1993).

On the market side, empirical evidence suggests that having a leading product in a therapy class is not a predictor of the likelihood of having a leading (in house originated) product in the next generation of therapies for that disease. Sutton (1998) analyzed the top 50 selling drugs in 1960, 1973, and 1986, focusing on the patterns of entry of new drugs. The analysis shows few instances where firms have been able to maintain their leading position in the submarket, pointing to a low degree of serial correlation in success. Rather, market shares are likely eroded by rivals, introducing the large majority of chemically related compounds that have followed the introduction of a top-selling product.

Evidence based on the US market dynamics shows that the reduction in the present discounted value of the innovator's return from between-patent competition, i.e. from new drugs introduced in the same therapeutic category, appears to be at least as large as the reduction from within-patent competition, induced by generic producers at patent expiration, and may be much larger (Lichtenberg and Philipson 2002).

All in all, entry and competition in R&D is higher in pharmaceuticals than in other sectors.

These features, coupled with ease of imitation of pharmaceutical compounds, make patents an important means for protecting innovations and, as a result, a good measure of the research effort in the pharmaceutical domain⁶.

⁶Results from the Carnegie Mellon Survey, about the nature and strength of the appropriability conditions in the US manufacturing sector, administered in 1994 (Cohen et al. 2000), show that the drug industry is the one where patents received the highest

We exploit the information provided by patents and patent citations, as a proxy for research efforts and knowledge diffusion, in order to characterize the dynamic nature of the innovation process in pharmaceuticals.

When using patent-based indicators for measuring technological change is important to recognize the fact that “the quality of the underlying innovation varies widely from patent to patent” (Scherer 1965), meaning that innovations differ substantially in terms of their technological and economic impact. The first empirical account of the heterogeneity in the private value of patents was based on information about the renewal fees of European patents (Pakes 1986, Schankerman and Pakes 1985, Schankerman and Pakes 1986), confirming that the distribution of patent value is highly skewed: a large amount of patents has a very low value, while the patents that provide important advance both in the technological and economic respects are quite few. Since then, various information regarding the patent documents have been used in the economic literature as a proxy for patent value, including the number of claims, the family size (i.e. the number of countries in which a patent application was filed on the innovation); the number of backward and forward citations; whether the patent was litigated and the years of renewal of the patent fees⁷.

We exploit the citation-value (both technological significance and economic impact) relationship largely investigated by the empirical literature on innovation and supported by survey evidence (Jaffe and Trajtenberg 2002, Harhoff, Narin, Scherer and Vopel 1999, Jaffe et al. 2000), and focus on the pattern of citations received by a patent, as a proxy for knowledge utilization and diffusion.

3 Data and Methods

The analysis builds on a comprehensive dataset on the innovative activity in the pharmaceutical industry, including R&D project level data, patents, citations, and collaborations.

An elaborate matching process has been employed to link together two

score as an effective mechanism for protecting property rights. In addition, 100% of respondents stated one of the reasons for patenting product innovations rely in preventing rivals from copying the innovation. Based on a different survey about the patenting activity of the largest European firms, Arundel and Kabla (1998) report a high propensity to patent in pharmaceuticals: the combined rate for both product and process innovations is higher than 50% in this industry.

⁷For a review see Lanjouw, Pakes and Putnam (1998); Hall, Jaffe and Trajtenberg (2001); Jaffe and Trajtenberg (2002).

different datasets related to the innovative activities in pharmaceutical, drawn from ATADB, maintained by ATA S.p.A. in Lucca, Italy.

The database contains information about all pharmaceutical and biotechnology patents⁸ granted by the USPTO since 1965, including their patent citations. Moreover, firm data at the level of specific R&D project worldwide in the last 25 years are available. ATADB tracks the development history of each R&D project, starting from patent application to the latest stage of drug development through preclinical and clinical development⁹. By exploiting the information about the patents protecting the compound the project data have been matched to patent data¹⁰, including the number of forward citations up to May 2004, the application date, and the name of the assignee.

The matching of the two sets proved to be a formidable, large-scale task, that tied up a great deal of our research efforts for a long time, providing us a unique dataset that monitors R&D activities of pharmaceutical and biotechnology firms from patenting to eventual commercialization of the protected compound¹¹.

Only patents associated to R&D projects of candidate drugs that were terminated either with a success, i.e. a product commercialised on the market, or with a failure, i.e. the project was discontinued due to the emergence of toxicological effects or to lack of effectiveness have been considered. The final database encompasses information about 2,000 R&D projects and their associated patents¹².

As a benchmark, we considered a sample of “matched” patents, selected on the basis of application year, publication year, and IPC classification characterizing the marketed and discontinued patents included in the analysis, sampling from biopharmaceutical patents with no information about preclinical or clinical development. By comparing the citation patterns of discontinued and marketed patents against the selected sample of matched patents, we will be able to ascertain the level of knowledge utilization and diffusion associated with each set and the related R&D competition dynamics.

⁸US patents in selected IPC and US classes are included in the database.

⁹In case of aborted projects, the time of discontinuation is reported.

¹⁰For projects listing a patent granted by a patent office other than the USPTO, we considered the US patent in the same family as the one listed in the database. In case no US patent is identified, the project is not considered in the analysis.

¹¹Old molecules and/or natural products, which do not have any associated patent have been omitted from the analysis.

¹²We refer to marketed (discontinued) patents as the patent associated to marketed (discontinued) R&D projects.

Since the focus is on the pharmaceutical industry, only citations from subsequent patents in the pharmaceutical domain have been taken into consideration. As usual, we distinguish self-citations from citations made by other companies. Citations made by others have been proved to be a good proxy for knowledge spillovers (Jaffe et al. 2000), whereas self-citations are considered to be indicators of the cumulative nature of the technology and a measure of the extent to which innovators are able to reap the benefits of their own research (Hall et al. 2001).

First, we present descriptive evidence distinguishing the citations received before and after the outcome of the associated R&D project becomes known.

Next, we will employ the double-exponential function to model the citation lag distribution for marketed and discontinued patents, against the average biopharmaceutical patents (i.e. the sample of “matched” patents). The model provides a flexible framework for studying the process underlying the generation of citations, where an exponential process by which knowledge diffuses is combined with a second exponential process by which knowledge become obsolete (Jaffe and Trajtenberg 1996, Caballero and Jaffe 1993). Following Jaffe and Trajtenberg (1996), we model the likelihood that a patent granted in year T will cite some particular patent granted in year t as:

$$p(t, T) = \alpha \exp[-\beta_1(T - t)](1 - \exp[-\beta_2(T - t)])$$

where α is linked to the overall likelihood of receiving a citation, whereas β_1 and β_2 are indicators of, respectively, the rate of obsolescence of knowledge (i.e., the rate at which new knowledge replace the existing one) and the rate of diffusion of the knowledge related to the invention protected by the patent. It is not possible to separately identify the three parameters in the model. We allow the parameters α and β_1 to vary as a function of attributes of both the cited and the citing patent (particularly, we distinguish marketed and discontinued patents from the sample of matched patents with no information about preclinical or clinical development), whereas β_2 is treated as a constant.

The analysis allows us to ascertain knowledge utilization and diffusion that is associated with marketed and discontinued patents with respect to the other patents issued within the biopharmaceutical domain.

Finally, we run a regression where the dependent variable is the number of citations received by our sample of patents adjusted on the basis of the estimated citation lag distribution, in order to reflect life-time citations. The estimation aims at identifying the factors that affect the importance of the

patent and of the associated innovation. We include among the regressors a set of dummy variables identifying cases where the patent is building on a previous failure or success and whether the patent is cited by subsequent successes/failures. Control variables for the characteristics of the technological classes (defined on the basis of the International Patent Classification), assignees and the patent itself are added.

The two sets of results will provide a clear picture of the dynamics underlying R&D competition in the biopharmaceutical domain. The estimation of the citation lag distribution function will allow us to show the dynamics associated with knowledge utilization and diffusion, whereas the regression analysis aims at disentangling the relationship between the productivity of knowledge and its sources.

4 Empirical Results

Figure 1 depicts the observed and estimated citation lag distribution functions for marketed and discontinued patents comparing them with patents with no information about preclinical and clinical development. The function represents the likelihood that any patent will be cited by the patents granted in the following years, as a function of time elapsed from the grant date (Jaffe and Trajtenberg 1996).

The x-axis in Figure 1 reports the citation lag, i.e. the difference between the citing and the cited patent grant year. On the y-axis the observed and estimated citation intensities are represented, defined as the ratio between the number of citations received and the potential number of citations.

The observed citation lag distribution is computed using the following formula:

$$p(t, T, o) = \frac{c(t, T, o)}{n(t, o) n(T)}$$

where t indicates the grant year of the cited patent, T is the grant year of the citing patent¹³, and o represents the outcome of the associated R&D project (marketed or discontinued vs. matched patents). The potential number of citation is given by the number of citations that would have been observed if all patents granted in year T would have cited all patents granted in year t with outcome o (marketed or discontinued), that is equal to the product of the number $n(T)$ of patents granted in the citing year and the number $n(t, o)$ of patents granted in the cited year with a known outcome o .

¹³The citation lag is given by $T - t$.

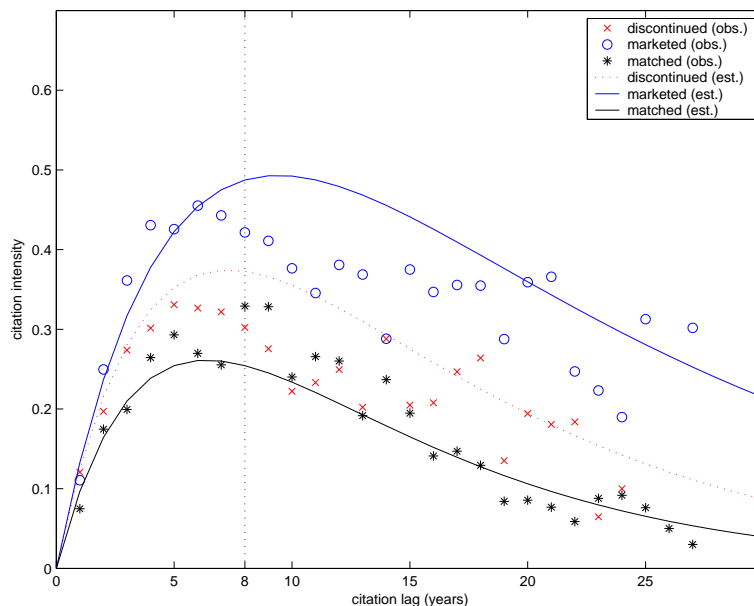


Figure 1: Observed and estimated average citation lag distribution

The estimated distributions are obtained from the following equation:

$$p(k, K) = \alpha(t, T, o) \exp[-\beta_1(o)(T - t)](1 - \exp[-\beta_2(T - t)]).$$

We assume that the grant year of the citing and the cited patent¹⁴ only affect α , while the outcome of the project affect both α and β_1 . This specification allows us to study how patents differ in the speed and extent to which existing knowledge is “picked up” in the case of failures and successes, as a proxy for the diffusion and utilization of the associated innovation. The results of the nonlinear least squares estimation¹⁵ are reported in Table 1 and the estimated densities are depicted in Figure 1.

Coherently with previous literature showing that the number of citations received by a patent is positively associated to its value¹⁶, citations turn out

¹⁴As in previous empirical literature dealing with this model, convergence problem forbids the estimation of the model where all the cited-year effects are considered. The problem is solved by introducing the cited-year effects considering 5-year time periods.

¹⁵Following previous empirical work we used a weighted method, where each observation is weighted by $(n_{to} n_T)^{1/2}$.

¹⁶See Trajtenberg (1990); Lanjouw and Schankerman (1999); Harhoff et al. (1999); Jaffe et al. (2000); Trajtenberg, Henderson and Jaffe (1997); Jaffe and Trajtenberg (2002).

	Model 1	Model 2
$\alpha^{discontinued}$	1.264** (0.182)	1.329** (0.111)
$\alpha^{marketed}$	1.309** (0.158)	1.511** (0.108)
β_1	0.107** (0.008)	0.084** (0.013)
$\beta_1^{discontinued}$	0.827** (0.135)	1.049** (0.108)
$\beta_1^{marketed}$	0.554** (0.087)	0.604** (0.077)
β_2	0.114** (0.014)	0.248** (0.042)
Cited year effects	no	yes
Citing year effects	no	yes
R-squared	0.686	0.858

* statistically significant at 5% level.

Table 1: Results of the double-exponential estimation, dependent variable: citation intensity

to be related to the outcome of the project. The observed and estimated distributions indicate that, on average, discontinued patents receive a number of citations that is lower than the number of citation received by patents associated to marketed projects (compare the values of the $\alpha^{discontinued}$ and $\alpha^{marketed}$ coefficients). However, both receive a higher number of citations than the average patent. The estimate of the α coefficients associated to discontinued and marketed patents are higher than 1, indicating that patents associated to preclinical or clinical development, irrespective of their outcome, are more likely to receive a citation than the average biopharmaceutical patents (taken as the base category). The analysis shows that there exists a value associated also with discontinued patents: even though the research associated with the patent will never reach the market, due to the emergence of toxicological problems or lack of effectiveness, the opened research trajectory is a source of information and insights for firms other than the original innovator.

The analysis of the estimates of β_1 reveals an important difference between marketed and discontinued compounds. The β_1 associated with associated with discontinued patents is very close to 1, pointing to no differences

between discontinued patents and the matched set of patents in terms of the rate of obsolescence of knowledge. On the contrary the β_1 associated with marketed patents is lower than 1, indicating that the knowledge embedded in patents protecting marketable compounds become obsolete less quickly than the other patents in the biopharmaceutical domain. The maximal citation frequency for discontinued patents is earlier in time than the maximal citation frequency of marketed patents.

The Figure also indicates the 8-year citation lag, which corresponds to the average length from patent application to termination of the project¹⁷.

Our estimates are coherent¹⁸ with the estimates of the Drugs and Medical sector presented by Hall et al. (2001). Moreover, an interesting patterns emerge in their results when comparing Drugs and Medical to other sectors. The citation lag distribution for this sector is more flat, whereas the citation lag distribution functions for the sectors of Computers and Communications, Electrical and Electronics, Chemical, and Mechanical have higher peak earlier in time. Knowledge in the Drugs and Medical sector diffuses less rapidly and takes a longer time to become obsolete. Important information about the protected compounds in terms of toxicological effects and effectiveness are revealed over time leading to a lengthier process of citation within this industry.

Furthermore, in order to analyze the patterns of technological competition, the role of information provided by science and R&D outcomes, we separately consider the pattern of citation (as a proxy for the diffusion and utilization of knowledge) before and after the outcome of the R&D project associated to the patent become known¹⁹, i.e. the R&D project is either discontinued or ends with the launch of a product on the market.

The first two columns of Table 2 report the share of citations received, respectively by discontinued and marketed patents after the project is terminated (either marketed or discontinued)²⁰. The Table reports the figure both for citations by other firms and self-citations.

As expected, marketed patents continue to receive citations by firms other than the original patentee also after the outcome of the project be-

¹⁷This is actually few months longer for marketed compounds, being equal to 7.8 years for discontinued R&D projects and to 8.3 years for marketed R&D projects.

¹⁸With respect to Hall et al. (2001) results, our estimated β_2 coefficient is lower. This might be explained by the fact that we only consider citations by institution other than the original assignee, which can require a longer time span with respect to self-citations to manifest.

¹⁹We compare the date of termination of the project with the application year of the citing patents.

²⁰100 represents the total number of observed citations.

Outcome	% citation after the outcome		avg. number of early citations	
	self-cits.	cits. by other	self-cits.	cits. by other
Discontinued patents	19.9	45.7	0.88	1.68
Marketed patents	43.1	63.3	0.87	1.95

Table 2: Share of citations received after the outcome of the projects, and early citations (within five years from patent application)

comes known, while discontinued patents received the largest share of citations before the time of their failure (more than 80 per cent).

On the one side, it is not surprising that successful compounds, leading to products that are commercialised on the market, continue to induce research (i.e. to attract citations) after the commercialization of the compound. Indeed, as it has been observed in the literature, after the introduction of a new product of major therapeutic value, the rivals of the innovating firm explore new lines of research trying to develop similar or related drugs (Sutton 1998). On the other side, it is of interest to understand the rationale behind the citations to discontinued compounds, still taking place after the outcome of the associated project has become known. Indeed, a share of discontinued patents still receives citations many years after the termination of the associated R&D project. Even though the compound under study is known to be toxic or ineffective for the targeted disease, subsequent research continues to build on the discontinued compound, as testified by citation from subsequent patents.

In most cases failures are not the subject of a publication, therefore very few information are available about the reasons behind the discontinuation of the research related to the compounds. The unique information in the public domain about the characteristics of the compound under development can be found in the patent(s).

By comparing the share of self-citations and the share of citations received by other companies, interesting conclusions can be drawn about the nature of the R&D competition in pharmaceuticals.

In the case of discontinued patents, about 80 per cent of self-citations is made before the compound is known to be a failure, while the share decreases to less than 60 per cent for marketed patents, i.e. a large part of patents associated to discontinued R&D projects are abandoned by the innovating firm after the properties of the associated compound are understood, but they still represent the ground for subsequent innovation by other companies.

We also took into consideration the number of citations in the early years after the application of the patent (within the first five years), before the termination of the project. The last two columns of Table 2 reports the average number of early citations received respectively by marketed and discontinued patents. Projects are taken into account only if not terminated, allowing us to compare the number of citations received by successes and failures when no certain information is available about their outcome.

No significant difference emerges from the comparison of the two distributions and their means²¹. This result coupled with previous findings suggests that differences in the pattern of citation between successes and failures are driven by post-outcome behavior, i.e. from the citations received by the patent after the launch of the associated product on the market²².

The disclosure of the information about the compound under study in patents and the advances in science sets the ground for a “race” for reaching the market, where competitors start exploring the new research arena pursuing parallel research trajectories even though the outcome is still highly uncertain. Competition on the R&D side in the pharmaceutical industry is substantial and firms entering the new research arena build both on future failures and successes.

The regression model presented in the following aims at adding new insights into the issue by looking at the relevance of the research building on failures and on successes, and of the research cited by future failures and successes.

The dependent variable is the (log) number of citation received during the life time of the patent, where the observed citation frequency has been adjusted using the estimated coefficient of the citation lag distribution function²³ (Table 1).

The independent variables are listed in Table 3 and aim at capturing the characteristics of the cited patent, of the IPC class and of the patenting firm. Moreover by taking into account whether R&D projects associated to backward and forward citations have been discontinued or have successfully reached the market, we aim at contributing to the discussion about the productivity of R&D spillovers (Levin and Reiss 1988).

²¹We performed both a standard t-test for mean comparison, and a Kolmogorov-Smirnov test. In both cases, the null hypothesis cannot be rejected at the usual level of significance.

²²Further research is needed to properly address this issue. Moreover, in order to distinguish “real” knowledge spillovers, it would be useful to distinguish the citations added by the patent examiner.

²³We consider the log of variable plus one in order to take into consideration patents receiving zero citations.

Variable	Description	Mean	σ
failure	Dummy equal to 1 if the associated project is a failure	0.34	0.48
success	Dummy equal to 1 if the associated project is marketed	0.14	0.34
ipc-nimp	Number of firms operating in the same IPC class	112.75	125.33
ipc-conc	Concentration of the IPC Class (Herfindhal index)	0.09	0.19
pt-selfc	Share of self-citations of the patent*	0.15	0.30
pt-orig	Index of originality of the patent*	0.42	0.37
pt-science	Science Index*	0.32	0.34
pt-importb	Importance of cited patents*	89.44	484.36
pt-timeb	Average time lag*	5.38	4.41
ass-coreec	Share of firm patent within the same technology class (IPC)	0.14	0.24
dbf	Dummy equal to 1 if the originating firm is a dedicated biotechnology company	0.14	0.35
pro	Dummy equal to 1 if the originator is a public research organization	0.13	0.33
bf	Dummy equal to 1 if the project-patent cites a previous failure	0.09	0.29
bs	Dummy equal to 1 if the project-patent cites a previous success	0.10	0.30
ff	Dummy equal to 1 if the project-patent is cited by a future failure	0.09	0.28
fs	Dummy equal to 1 if the project-patent is cited by a future success	0.06	0.25

* defined as in Trajtenberg et al. (1997).

Table 3: Description of the variables

Few patents cited by or citing our sample of patents are associated to a R&D project with a known outcome. Particularly, we identify 9 per cent of patents building on a previous failure, and 10 per cent of patents building on a previous success. On the other side, 6 per cent of patents are cited by a future success, and 9 per cent by a future failure. This reflects the fact that pharmaceutical and biotechnology firms screen thousands of compounds but very few enter into preclinical and clinical stages of development.

As far as the characteristics of the technological class of the patent, we consider the number of firms active in the IPC class, and the Herfindahl index of concentration computed at the technology class level on the basis of patent counts.

Patents characteristics are measured using the indicators developed by Trajtenberg et al. (1997) on the basis of backward citations. We consider the share of self-citations in the patents (*pt-selfc*) that measures the extent to which benefits from research antecedents are appropriated by the firm and help in understanding whether the patent belongs to a research trajectory strongly rooted within the company. The index of originality of the patent (*pt-orig*) measures the breadth of its technological roots, whereas the importance of the previous patents cited by the patent under investigation is measured by *pt-importb* which takes into account the number of backward citations in the patents and the number of citations they receive. The importance of scientific sources with respect to technological ones within the patent is captured by *pt-science*, which is the ratio between the non-patent references and the total number of references (previous patents or previous scientific literature) listed in the patent. The closer to 1, the larger the scientific underpinnings of the research, relying more heavily on the scientific literature rather than on previous patents. Finally *pt-timeb* measures the time distance between the citing and the cited patents. The higher *pt-timeb*, the older the sources the patent builds upon.

As compared to the descriptive statistics reported in Trajtenberg et al. (1997), no difference emerges with respect to the value of *pt-selfc*. On the contrary, the average value of *pt-timeb* in our sample is lower, indicating younger sources for our sample of patents, whereas the values of *pt-orig*, *pt-science*, and *pt-importb* are higher. One important difference with the sample in Trajtenberg et al. (1997) relies in the fact that we only consider pharmaceutical patents, and citations are counted only within the pharmaceutical technological classes.

As far as the characteristics of the patent assignee are concerned, we take into account the share of firm patent within the same technology class (IPC), and two dummy variables indicating whether the patentee is a dedi-

	Model 1	Model 2	Model 3	Model 4
failure	0.556 (0.056)**	0.552 (0.056)**	0.421 (0.054)**	0.357 (0.059)**
success	1.221 (0.075)**	1.204 (0.076)**	0.985 (0.074)**	1.001 (0.082)**
ipc-conc	-0.337 (0.144)**	-0.330 (0.144)**	-0.255 (0.138)*	-0.244 (0.138)**
ipc-nimp	-.5E-3 (.2E-3)**	-.5E-3 (.2E-3)**	-.6E-3 (.2E-3)**	-.5E-3 (.2E-3)**
pt-selfc	-0.222 (0.087)**	-0.245 (0.088)**	-0.293 (0.084)**	-0.279 (0.084)**
pt-orig	-0.405 (0.080)**	-0.379 (0.081)**	-0.353 (0.078)**	-0.358 (0.078)**
pt-science	0.697 (0.090)**	0.685 (0.090)**	0.660 (0.086)**	0.665 (0.086)**
pt-importb	.4E-3 (.5E-5)**	.4E-3 (.5E-5)**	.3E-3 (.5E-5)**	.4E-3 (.5E-5)**
pt-timeb	-0.004 (0.006)	-0.007 (0.007)	-0.005 (0.006)	-0.005 (0.006)
ass-corec	0.172 (0.108)	0.178 (0.108)	0.230 (0.104)*	0.227 (0.104)**
dummydbf	0.548 (0.079)**	0.554 (0.079)**	0.525 (0.076)**	0.536 (0.076)**
dummypro	0.384 (0.078)**	0.389 (0.078)**	0.362 (0.075)**	0.360 (0.075)**
bf		0.058 (0.088)	0.008 (0.085)	-0.307 (0.125)**
bs		0.186 (0.084)**	.2E-3 (0.083)	0.068 (0.124)
ff			1.093 (0.084)**	1.097 (0.084)**
fs			0.913 (0.101)**	0.908 (0.101)**
bf*fail.				0.666 (0.176)**
bf*succ.				0.131 (0.295)
bs*fail.				-0.052 (0.183)
bs*succ.				-0.208 (0.202)
Constant	1.180 (0.156)**	1.168 (0.156)**	0.960 (0.151)**	0.972 (0.150)**
Log Likel.	-6886.06	-6883.37	-6752.32	-6744.40

Table 4: Regression results. Dependent variable: $\ln(\text{number of adjusted forward citations} + 1)$. Number of observation: 4,285.

cated biotechnology company (dfb) or a public research organization (pro). The largest share of patents in our sample are assigned to pharmaceutical companies: 14 per cent of patents are assigned to dbf, and 13 per cent in the case of pro.

Results of the estimation of a censored regression model are reported in Table 4. Cited year dummies are included in all the specifications. The estimation procedure has been preferred to simple regression due to the high incidence of patents receiving zero citations (27.28%).

Coherently with previous results and with the empirical literature supporting the use of citations received by a patent as a proxy for its value both in economic and technological terms (Trajtenberg 1990, Lanjouw and Schankerman 1999, Harhoff et al. 1999, Jaffe et al. 2000, Trajtenberg et al.

1997, Jaffe and Trajtenberg 2002), we find that patents associated to marketed R&D projects receive a higher number of citations than patents with no preclinical or clinical information. However, also discontinued patents receive a higher number of citations than our sample of matched patents (taken as the benchmark category), even though the estimated coefficient of the failure dummy is lower than the estimated coefficient of the success dummy. Coherently with results reported in Table 1, discontinued patents receive a lower number of citations than marketed patents.

As far as the share of self-citations, the variable has a negative impact on the number of forward citations by firms other than the original assignee, supporting the claim that self-citations are indicative of the level of appropriability of research efforts. Being strongly rooted within the technological domain of the patent assignee significantly affects the number of citations subsequently received by firms other than the original innovator, i.e. the level of knowledge utilization.

The estimated coefficients of *pt-orig* shows that patents with sparse technological roots receive a lower number of citations by other firms. These are likely patents within narrow fields of application, therefore being relevant only to the firms and institutions working within the same technological domain.

Patents with predominance of scientific sources over technological ones contribute more heavily to subsequent research. This is not surprising within the biopharmaceutical domain, characterized by a strong link of innovative activity to its scientific underpinnings. Also patents building on an important (in the sense of highly cited) knowledge base are more often subsequently cited.

The estimated coefficients of *pt-timeb*, the average age of the sources the patent builds upon, is negative but not statistically significant. Competition in the pharmaceutical domain is substantial and the relevant knowledge base is rapidly evolving, pushing the innovating firms need to rely on the most recent advances and discoveries. Patents relying upon older knowledge bases have a low innovative content, and will be less often the basis of subsequent research. However, the age of the backward citations doesn't seem to affect significantly the number of forward patent references.

As far as the characteristics of the patent assignee are concerned patents by *dbf* and *pro* receive on average a higher number of citations. The former result is consistent with Hall, Jaffe and Trajtenberg (2000) who find that in the pharmaceutical sector smaller biotechnology firms are more likely to average a higher citation rate. As a tentative explanation for this phenomena, we propose that this is due the growing division of innovative labor and the

wide network of collaborations among the different actors involved in the drug development process that has come to characterize the pharmaceutical industry (Arora and Gambardella 1994, Powell et al. 1996, Orsenigo et al. 2001). The small biotechnology firms are highly specialized in the early stages of drug development, but they lack the resources and capabilities that are needed for the large clinical trials, therefore they are more likely to license-out their compounds to the large pharmaceutical companies with significant expertise with clinical trials. This is also true for public research organizations, strongly oriented toward basic science and the early stages of the innovation process in pharmaceuticals. As a result, dbf and pro patents are more likely to be the basis of subsequent research by firms other than the original innovators, particularly they will be the subject of the research undertaken by the company that licensed-in the compound and continues research around it. If the research undertaken by the licensee gives rise to new technological opportunities or compounds, the new innovation will certainly cite the licensed patent, therefore increasing the number of citations received by competitors to dbf and pro patents.

The characteristics of the IPC class of the patent also exert a significant effect on the number of subsequent citations received by the patent. Patents in classes that are highly concentrated receive on average a higher number of citations. This result might be driven by the low incentives of firms other than the original innovators in pursuing research in classes where concentration on the technological side is high, i.e. the technological competences are bundled within a low number of firms. On the contrary, a higher number of firms in the technological class leads to a lower number of citations, *ceteris paribus*. Patents in crowded technological fields receive, on average, a lower number of citations, due to the larger number of different research trajectories pursued by different firms that can be the source of knowledge in subsequent developments.

Looking at the relevance of previous and future failures or success citing or cited by the patent, we note that patents that are subsequently cited by future failures and successes receive on average a higher number of citations. On the contrary, previous successes and failures do not have any statistically significant effect on the relevance of the patent for subsequent research²⁴, pointing to low cumulativeness of research and high uncertainty in the pharmaceutical domain. Building on a previous success doesn't assure to reach

²⁴The estimated coefficient of *bs* (a dummy variable equal to one if the project-patent cites a previous success) is positive and significant in Model 2, but the result is not robust to the inclusion of the *ff* and *fs* dummies.

high levels of R&D productivity.

However, looking at the projects that entered into preclinical or clinical trials reveals that patents citing previous failures receive a higher number of citations than the base category. The composite effect is statistically significant for failed R&D project patents. This pattern might be the result of the process of trial-and-error at work within the pharmaceutical domain. The estimated coefficients support a role of failures in spurring technological competition in the pharmaceutical domain, being the basis of future research by other firms, which seems to reach higher levels of efficiency in pursuing the research, at least in terms of diffusion and utilization (as proxied by the number of received citations).

Overall the analysis reveals that the information contained in patents represent an important source of information for monitoring the R&D activities undertaken by the competitors and provide a spur to innovative efforts by other firms in related fields or in the same area of application of the original patent.

5 Discussion and Policy Implications

This paper has looked at the nature of R&D competition in the pharmaceutical industry, a unique framework for studying issues related to innovation and innovative activities. Besides the importance of patents as a means for appropriating returns to R&D, as evidenced in surveys of firms in the manufacturing sectors in the US and Europe (Arundel and Kabla 1998, Cohen, Goto, Nagata, Nelson and Walsh 2002), a number of evolutionary trends has profoundly shaped the organization of innovative activities within the industry.

The industry has come to be the archetype of the “science-based” sector, where advances in basic knowledge about bacterial, animal and human processes provide a deeper understanding of the molecular and biochemical roots of specific diseases processes and of the mechanisms of pharmacological action of known and new substances, guiding the innovative activities of the actors involved in drug development. The advances in genomics, gene sequencing, transgenic animals have provided the industry a huge number of novel biological targets thought to be relevant to a vast array of diseases. As a result, firms likely pursue parallel research trajectories searching for compounds with binding properties around the same target.

Against this background, the paper has explored the learning process of the actors involved in the pharmaceutical domain. Differently from previous

studies, we also take into account the role of research failures in providing the ground for subsequent innovation.

Patent protection forbid direct imitation of the compound (or process), nonetheless rival firms may search around the original molecule and find a patentable variant that offer some advantages. Patent citation data allow to track subsequent developments.

Coherently with previous literature showing the existence of a relationship between patent citation and (private and social) patent value, marketed patents receive a higher number of citations, and for a longer time span. No difference emerges between marketed and discontinued patents in the early stages of development, where the research outcome is still unknown. Differences in the citation lag distribution is driven by the post-outcome behavior, i.e. from the citations received after the successful product commercialization.

However, our empirical analysis shows the existence of a set of discontinued compounds whose patent receives a high number of citations, also after the termination of the associated R&D project, i.e. they provide the basis for subsequent innovations, eventually also leading to compounds that successfully reach the market. This suggests the existence of a social value associated to discontinued project and to the diffusion of the associated information, in terms of new (and better) research trajectories exploring new therapies for treating a disease. Moreover the regression analysis shows that research building on rival failures has a wider impact on subsequent research, as measured by the number of life-time citations received by the associated patent.

The information disclosed through patents leads to an expansion of the knowledge frontier, that stimulate further R&D effort, both in terms of new patents and new firms entering the research arena. It may well happen that firms other than the original innovator are the first to reach the market, being able to pursue more effectively the new line of research.

In this perspective, the discussion about patent scope becomes crucial in this industry, where research is highly cumulative in nature and firms enjoy knowledge spillovers spanning from internal and, to some extent, from external R&D projects, pointing to a trade-off that cannot be easily resolved. This poses problems for the optimal design of patent law²⁵. On the one side, it is necessary to fully reward early innovators for the technological foundation they provide to later innovators, but also later innovators should

²⁵See Scotchmer (1991) for a detailed discussion of the optimal patent scheme in the case of cumulative knowledge.

be rewarded adequately for their improvements and new products.

Too narrow patents would be ineffective as incentives to R&D (the main function of the patent system), whereas too broad patents would be an obstacle to the development of parallel research trajectories, which might be improved with respect to their predecessors in terms of side effects or delivery method. In addition, as shown in Scotchmer (1991), too broad patents might inefficiently inflate incentives for the first innovator, which might not be capable of efficiently pursuing all subsequent lines of research.

References

- Alcacer, J. and Gittelman, M.: 2004, How do I know what you know? Patent examiners and the generation of patent citations, *New York University, mimeo*.
- Arora, A. and Gambardella, A.: 1994, The changing technology of technological change: general and abstract knowledge and the division of innovative labour, *Research Policy* **23**(5), 523–532.
- Arundel, A. and Kabla, I.: 1998, What percentage of innovations are patented? empirical estimates for European firms, *Research Policy* **27**(2), 127–141.
- Breschi, S., Malerba, F. and Orsenigo, L.: 2000, Technological Regimes and Schumpeterian Patterns of Innovation, *The Economic Journal* **110**(463), 388–410.
- Caballero, R. and Jaffe, A.: 1993, How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth.
- Cohen, W., Goto, A., Nagata, A., Nelson, R. and Walsh, J.: 2002, R&D Spillovers, Patents and the Incentives to Innovate in Japan and the United States, *Research Policy* **31**(8), 1349–1367.
- Cohen, W., Nelson, R. and Walsh, J.: 2000, Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not).
- DiMasi, J., Hansen, R. and Grabowski, H.: 2003, The price of innovation: new estimates of drug development costs, *Journal of Health Economics* **22**(2), 151–185.

- Dosi, G.: 1988, Sources, Procedures, and Microeconomic Effects of Innovation, *Journal of Economic Literature* **26**(3), 1120–1171.
- Hall, B., Jaffe, A. and Trajtenberg, M.: 2000, Market Value and Patent Citations: A First Look.
- Hall, B., Jaffe, A. and Trajtenberg, M.: 2001, The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools.
- Harhoff, D., Narin, F., Scherer, F. and Vopel, K.: 1999, Citation Frequency and the Value of Patented Inventions, *The Review of Economics and Statistics* **81**(3), 511–515.
- Henderson, R. and Cockburn, I.: 1994, Racing to Invest? The dynamics of competition in ethical drug discovery, *Journal of Economics and Management Strategy* **3**(3).
- Henderson, R. and Cockburn, I.: 1996, Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery, *Rand Journal of Economics* **27**(1), 32–59.
- Jaffe, A. and Trajtenberg, M.: 1996, Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries.
- Jaffe, A. and Trajtenberg, M.: 2002, *Patents, citations and innovations: a window on the knowledge economy*, MIT Press.
- Jaffe, A., Trajtenberg, M. and Fogarty, M.: 2000, The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees.
- Kling, J.: 1998, From Hypertension to Angina to Viagra, *Modern Drug Discovery* **1**(2), 31–38.
- Lanjouw, J., Pakes, A. and Putnam, J.: 1998, How To Count Patents and Value Intellectual Property: The Uses Of Patent Renewal and Application Data, *The Journal of Industrial Economics* **46**(4), 0022–1821.
- Lanjouw, J. and Schankerman, M.: 1999, The Quality of Ideas: Measuring Innovation with Multiple Indicators.
- Levin, R. and Reiss, P.: 1988, Cost-Reducing and Demand-Creating R&D with Spillovers, *The RAND Journal of Economics* **19**(4), 538–556.

- Lichtenberg, F. and Philipson, T.: 2002, The Dual Effects of Intellectual Property Regulations: Within- and Between-Patent Competition in the U. S. Pharmaceuticals Industry, *The Journal of Law and Economics* **45**, 643–672.
- Malerba, F. and Orsenigo, L.: 1993, Technological regimes and firm behavior, *Industrial and Corporate Change* **2**(1), 45–71.
- Narin, F., Smith jr., V. and Albert, M.: 1993, What patents tell you about your competition, *ChemTech* **February**, 52–59.
- Nelson, R. and Winter, S.: 1982, *An evolutionary theory of economic change*, Belknap Press.
- Oprea, T., Davis, A., Teague, S. and Leeson, P.: 2001, Is There a Difference between Leads and Drugs? A Historical Perspective, *Journal of Chemical Information and Computer Sciences* **41**(5), 1308–1315.
- Orsenigo, L., Pammolli, F. and Riccaboni, M.: 2001, Technological change and network dynamics: lessons from the pharmaceutical industry, *Research Policy* **30**(3), 485–508.
- Pakes, A.: 1986, Patents as Options: Some Estimates of the Value of Holding European Patent Stocks, *Econometrica* **54**(4), 755–784.
- Pavitt, K.: 1984, Sectoral patterns of technical change: towards a taxonomy and a theory, *Research Policy* **13**(6), 343–373.
- Powell, W., Koput, K. and Smith-Doerr, L.: 1996, Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology., *Administrative Science Quarterly* **41**(1).
- Schankerman, M. and Pakes, A.: 1985, The Rate of Obsolescence and the Distribution of Patent Values: Some Evidence from European Patent Renewals, *Revue Economique* **36**(5), 917–941.
- Schankerman, M. and Pakes, A.: 1986, Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period, *The Economic Journal* **96**(384), 1052–1076.
- Scherer, F.: 1965, Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions, *The American Economic Review* **55**(5), 1097–1125.

- Sutton, J.: 1998, *Technology and market structure: theory and history*, The MIT Press Cambridge, Mass.
- Trajtenberg, M.: 1990, A Penny for Your Quotes: Patent Citations and the Value of Innovations, *The RAND Journal of Economics* **21**(1), 172–187.
- Trajtenberg, M., Henderson, R. and Jaffe, A.: 1997, University vs. Corporate Patents: A Window on the Basicness of Innovations, *Economics of Innovation and New Technology* **5**(1), 19–50.
- Trenter, M.: 1999, From Test Tube to Patient: Improving Health Through Human Drugs, *Center for Drug Evaluation and Research Special Report. US Food and Drug Administration, Rockville, Md .*
- Winter, S.: 1994, Schumpeterian Competition in Alternative Technological Regimes, *Journal of Economic Behavior and Organization* **5**, 287–320.