



DRUID
ASIA

Paper to be presented at
DRUID-Asia, Singapore, February 23-25, 2016
Co-organized by DRUID, NUS Business School and SMU - Lee Kong Chian
School of Business.

What do patent-based measures tell us about product commercialization? Evidence from the pharmaceutical industry

Stefan Wagner
ESMT
Strategy and Innovation
stefan.wagner@esmt.org

Simon Wakeman

Strategy and Innovation
simon_wakeman@yahoo.co.nz

Abstract

Patent-based measures are frequently used as indicators in empirical research on innovation and technological change. Currently, there is little evidence as to what extent patent-based indicators relate to product market outcomes. Using a unique dataset that links outcomes from product commercialization in the pharmaceutical industry with detailed patent data, we relate patent-based indicators that capture either an invention's value or the uncertainty surrounding the patenting process to the outcomes of the product development process. Our findings suggest that the speed of commercialization increases with value but reduces with uncertainty. Using a variety of alternative indicators we derive implications for the use and the proper interpretation of individual measures. Moreover, our study has broader implications as it highlights the detrimental effect of uncertainty on the speed of innovation.

What do patent-based measures tell us about product commercialization? Evidence from the pharmaceutical industry

August 2015

Abstract

Patent-based measures are frequently used as indicators in empirical research on innovation and technological change. Currently, there is little evidence as to what extent patent-based indicators relate to product market outcomes. Using a unique dataset that links outcomes from product commercialization in the pharmaceutical industry with detailed patent data, we relate patent-based indicators that capture either an invention's value or the uncertainty surrounding the patenting process to the outcomes of the product development process. Our findings suggest that the speed of commercialization increases with value but reduces with uncertainty. Using a variety of alternative indicators we derive implications for the use and the proper interpretation of individual measures. Moreover, our study has broader implications as it highlights the detrimental effect of uncertainty on the speed of innovation.

Keywords: Patent indicators, patent system, product commercialization, pharmaceutical industry, drug development

1. Introduction

Patent-based measures are frequently used in empirical research on innovation and technological change and have become increasingly popular in diverse topics such as studies of firms' strategies and organizational choices as well as labor mobility or team performance. Early work primarily relied on simple patent counts as a measure of innovation output (see Griliches 1990 for a survey). More recently, the availability of comprehensive micro-level data has enabled the construction of more refined indicators aiming at characterizing the strength of the protective scope of a patent as well as describing the underlying inventions on a large scale. These indicators are derived from information contained in publicly available patent documents such as patent references, technology classifications or inventors involved. Refined patent-based indicators have been widely applied to study the outcomes and the functioning of the patent system itself (see Hall and Harhoff 2012 for a recent survey). Going beyond the patent system itself, patent indicators are increasingly used to study firm-level phenomena such as R&D productivity (Jaffe 1986), firm survival (Malerba and Orsenigo 1999; Nerkar and Shane 2003; Wagner & Cockburn 2010), investments in young companies (Sorenson and Stuart 2001; Cockburn and MacGarvie 2009) or entry in industries (Cockburn and MacGarvie, 2011), and alliance formation (Mowery, Oxley, and Silverman 1996; Stuart 1998).

As the use of these indicators increases, it is important to establish that these patent-based measures actually reflect the more general facts that they are implicitly claimed to represent. Having a clear understanding of the relationship between these indicators and the underlying (mostly unobserved) phenomena, such as an invention's value or the strength of patent protection, that they purport to measure is important not only for their construction but also for a meaningful interpretation of their effects. Much work has therefore been invested in the validation of patent indicators. First, the correlation between the value of an invention and the

number of citations a patent receives by subsequent applications has been clearly established (see Trajtenberg 1990; Harhoff, Narin, Scherer, and Vopel 1999; Gambardella, Harhoff and Verspagen 2008; Hall, Jaffe, and Trajtenberg 2005). Second, it has been highlighted that patent indicators are informative regarding the uncertainties surrounding the patenting process. These uncertainties pertain to the timing, the scope, and the legal stability of patent protection. Patent indicators have been linked to uncertain outcomes such as patent grant, post-grant validity challenges or the occurrence of licensing (Gans, Hsu, and Stern 2008; Harhoff and Reitzig 2004; Harhoff and Wagner 2009; Lanjouw and Schankerman 2001; Régibeau and Rockett 2010).

Despite these validation efforts, much less is known about which and to what extent patent indicators convey valuable information with regard to an invention's likelihood of commercialization that is the most immediate product-market outcome of innovative activities possible. Since patent indicators are often used to describe firms' competitive situation in the product market, it is of the utmost importance to establish a clear understanding of how to interpret the results when these indicators are used to derive conclusions that reach beyond the patent system. In this paper we add new insights into this important question by investigating how frequently used patent indicators associated with value and uncertainty affect actual outcomes from product commercialization processes. The study is rooted in the pharmaceutical industry for two reasons: First, obtaining a clear and unencumbered patent position is essential for commercializing an innovation in the pharmaceutical industry (Cohen, Nelson, and Walsh 2000). Patent rights in this industry are generally strong, which should allow for an easier detection of correlations between patent indicators and outcomes of the product commercialization process. Second, the pharmaceutical industry is classified as a 'discrete' industry where a relatively small number of patents (often only one patent) can protect an entire product (Cohen, Nelson, and Walsh

2000). This allows us to clearly link product development projects to patents protecting the underlying invention. In particular, we exploit a novel dataset that combines detailed information regarding the commercialization of 5,293 pharmaceutical development projects with fine-grained information on the underlying patents. Such a research design would not be possible in complex technologies where a large number of patents are associated with a single product.

Overall, our findings show that the effect of indicators related to an invention's value and the effect of indicators reflecting uncertainty regarding the scope and strength of patent protection are in line with expectations. We find that value-related measures such as the number of different patents surrounding a commercialization project, the number of countries in which primary patents have been applied for (family size), and different citation-based measures are all related to increased hazards and a higher speed of product commercialization. Further, our data reveal that uncertainty has a significant effect on the hazard of commercialization: Once a patent has been granted, and hence the scope of the IP protection of a potential drug has been clearly delineated, we observe that firms significantly speed up the commercialization process. This indicates that uncertainty regarding patent protection slows down commercialization.

Our contribution is twofold: First, to the best of our knowledge, our study is the first attempt to combine fine-grained patent indicators with product market-outcomes on a large scale. Based on this unique research setting, we are able to derive statements about the extent to which patent-based indicators are informative about outcomes beyond the patent system itself. In particular, we distinguish between indicators that are based on endogenous applicant behavior such as family size and measures that lie beyond the influence of the patent applicant such as citation-based measures. Our findings suggest that indicators which depend on applicant behavior are more informative. Second, and equally important, our findings bear

relevance for the design of patent systems. While previous research has scrutinized the effect of patenting on follow-on innovation (Sampat and Williams 2014; Williams 2013) and the effect of patent terms on the direction of R&D (Budish, Roin, and Williams 2013), we derive insights regarding the effect of uncertainty on the speed of commercialization. Once uncertainty has been resolved the hazard of product commercialization increases significantly. This finding is in line with related work analyzing the hazard of licensing (Gans, Hsu, and Stern 2008) and has broader implications regarding the effect of patent grant delays on the speed of innovation.

2. Pharmaceutical product commercialization and patent-based indicators

2.1 Product commercialization in the pharmaceutical industry

The pharmaceutical industry is characterized by a high R&D intensity and developing new drugs is expensive. Estimates of the average cost of development generally exceed US \$800 million (DiMasi, Hansen, and Grabowski 2003; Adams and Brantner 2006) and research-active pharmaceutical companies (originator companies) spend about 17 percent of their revenues from prescription drugs on R&D (European Commission 2009).¹ The commercialization process of inventions in this industry can be divided into three different phases: (i) the pre-launch period where R&D, clinical trials, and clinical tests take place, (ii) the marketing and sales period during which the originator company sells its product under exclusivity usually derived from patents, and (iii) a post-exclusivity period where competition by generic companies copying the initial invention is possible (European Commission 2009).

¹ Note that 1.5% of the revenues made from prescription drugs is spent on basic research while the remaining fraction is spent on clinical trials and tests (European Commission 2009).

In this paper, we specifically focus on the pre-launch period, in which the duration between the invention of a compound and its market launch, which is of utmost importance for companies commercializing pharmaceutical inventions. Initial patents relating to a drug are typically filed during the basic R&D stage and have a fixed duration of 20 years from the original filing date (known as the “priority date”). Hence, the length of the period during which a drug can be sold by the originator company under an exclusivity granted by the primary patent right is directly determined by how much time lapses between the filing of the patent and the market launch of a product. The shorter the duration of the pre-launch stage, the more time an originator company has to enjoy the exclusivity provided by patent protection and thereby recoup the investments made during drug discovery and testing without facing competition from companies selling drugs based on the same substance (Budish, Roin, and Williams 2013).

The pre-launch phase contains the search for molecular targets associated with a disease and the identification of novel pharmaceutically-active substances that interact with the target. Once a substance has been identified, various preclinical and clinical tests are carried out to ascertain the toxicity and efficacy of the new molecule. Finally, national regulators have to approve the drug before it can be sold on the market. The overall duration of the pre-launch period is of significant length. Sternitzke (2010) reports an average duration of about 11.5 years, while the European Commission (2009) reports a shorter period of only 8.6 years for a selection of 144 substances. After the initial discovery of pharmaceutically-active substances companies will usually file first patent applications related to the active molecules themselves.² Depending on the actual duration until a substance is first sold on the market

² These applications, and the resulting patents, are generally referred to as “primary patents” because they relate to the first patents for the active molecules.

originator companies are generally left with a period of eight to 12 years during which they enjoy exclusivity arising from a granted patent right.

There are various factors affecting the speed of product commercialization once a product candidate has been discovered. Most importantly, the duration of obligatory clinical trials considerably depends on characteristics exogenous to the firm such as the mortality rates of the disease to be treated.³ Budish, Roin, and Williams (2013) highlight that these characteristics not only affect the speed of commercialization but, due to the 20-year fixed term of protection, they can also distort firms' R&D decisions. Firms favor R&D in areas with higher mortality rates over areas where mortality is lower as they can complete obligatory clinical tests faster in the former. However, the social value of R&D often is higher in areas where clinical trials take longer (such as disease prevention) and Budish, Roin and Williams (2013) argue that the distortion might therefore be welfare destroying. In addition to exogenous factors, the speed of product commercialization will also depend on the firm behavior manifested in the effort (as well as the money) it invests in the commercialization of a drug. Companies can be expected to put more effort into projects with a (perceived) high private value and bring these to market faster than low-value projects. Private value, in turn, will depend on the value of the underlying invention and the possibility of capitalizing on that value, which strongly depends on the patent protection available (Sternitzke 2010, 2013).

In the following, we relate patent-based measures of (i) an invention's value, and (ii) the uncertainty regarding the appropriation of that value to progress through the different stages of the commercialization funnel. Although prior literature clearly establishes a relationship between a number of patent indicators and the value of inventions in general, it also shows

³ In our regressions we control for the ATC (Anatomical Therapeutic Class) of a drug to capture these differences.

that these measures explain only a small proportion of the variation in the value (Gambardella et al. 2008). While we might reasonably assume that value indicators are positively related to the technological aspects (i.e., whether the invention enters clinical trials), it is an open question as to whether they also provide any information on the likelihood that the invention will progress through clinical trials. Similarly, it is worth examining how measures describing uncertainty surrounding the appropriability of an invention's value influence the speed of commercialization in our context. Based on the argument above we expect value-related measures to be associated with higher commercialization hazards while indicators of the uncertainty of patent protection should slow down commercialization – at least until this uncertainty has been resolved over time.

2.2 Patent-based indicators

(a) Indicators related to value

The value of an invention depends on both its technological quality and the economic importance of the problem it solves. A large volume of literature has used a range of patent-based measures to characterize these aspects. It is beyond the scope of this paper to provide a full survey of patent indicators and their use; we restrict ourselves to highlighting the most important stylized facts.

The most widely used patent-based variable used to proxy the value of the underlying invention is the number of citations the focal patent receives from subsequent patents (Trajtenberg 1990; Harhoff et al. 1999), commonly referred to as “forward citations.” Correlating forward citations with responses from surveys of patentees with regard to the value they actually extracted from specific patents (Harhoff et al. 1999; Harhoff, Scherer and Vopel 2003; Gambardella et al. 2008) and to other measures of value such as stock market valuations measured in Tobin's q (Bessen 2009; Hall, Jaffe, and Trajtenberg 2005) or

indicators of public surplus (Trajtenberg 1990) reveals a significant correlation between the number of citations a patent receives and an invention's value.

Despite the well-documented correlation between forward citations and an invention's value the use of forward citations is not free of problems. Citations can either originate from patent applications of the same applicant (self-citations) or from patent applications of other applicants (usually called citations). Self-citations might be related to private value while citations reflect the technological value of an invention for a technological field. Hall, Jaffe, and Trajtenberg (2005) have shown that self-citations explain a larger fraction of variance in a focal firm's valuation than citations. Moreover, patent references that are the basis of citation-based measures can originate from the patent examiner or the patent applicant, depending on the institutional setting, further aggravating a clear interpretation (see Gittelman 2008a and 2008b and Moser, Ohmstedt and Rhode 2013 for a detailed discussion). Another disadvantage of using citation-based indicators of patent value is that they are available only with a considerable lag as citations only arrive over time. It is common to use the number of citations that a patent has received within five years after its application.⁴ To summarize, patent citations are frequently used but they are available only ex post and are at best a noisy measure of patent value. In the context of the commercialization of pharmaceutical inventions, we expect both the number of self-citations and the number of citations to be positively correlated with the speed of product commercialization. Firms should – ceteris paribus – be willing to invest more effort in the commercialization of inventions of a higher value.

⁴ The number of years a patent has been renewed has also been used as a value indicator (Pakes and Schankerman 1984; Pakes 1986; Schankerman and Pakes 1986; Lanjouw, Pakes, and Putnam 1998). To observe renewals, however, one has to wait even longer than for citation measures and information on the full renewal history is only available 20 years after their application.

While citations often arrive only years after the patent filing, alternative value indicators are available in earlier stages of the patenting process. They include the number of countries in which patent protection is sought (often called “family size”, Putnam 1996; van Pottelsberghe de la Potterie and van Zeebroeck 2008) and whether a patent was applied for under the rules of the Patent Cooperation Treaty (PCT) (Harhoff and Reitzig 2004; Harhoff and Wagner 2009). A patent’s family size captures the number of jurisdictions in which patent protection for a single invention has been sought while a PCT application opens up the opportunity to extend patent protection globally in a streamlined process. The PCT provides a unified procedure for filing patent applications to protect inventions in up to 147 contracting states within 30 months of the filing date. PCT applications allow the expansion of patent protection to a large number of countries without incurring the full costs and complexity of a national application path and hence have greater option value. It allows applicants to delay the choice of countries for which they designate the application for up to 30 months after the priority date compared to only 12 months for national or regional patent applications. The expansion of patent protection to different countries is costly and applicants are willing to incur these extra costs, particularly for valuable inventions. Therefore, we expect both measures – family size and PCT applications – to be related to higher commercialization hazards. It needs to be pointed out, however, that unlike patent citations (which are at least partially exogenous as they originate from other patent applicants) these indicators are endogenous: Firms will opt for a larger international patent protection only if the perceived private value of an invention is high.

While not a patent-level indicator, the total number of different patents surrounding a commercialization project can also be interpreted as an indicator reflecting a company’s assessment of a project’s value. In addition to the primary patent protecting the active compound underlying a product companies often file further patent applications for different

aspects of the active molecules. These are usually related to dosage forms (e.g., tablets, capsules or injection solutions) or for particular pharmaceutical formulations (mixtures of active agents and other substances which promote the activity of the medicine) – and are referred to as “secondary patents” (European Commission 2009). Secondary patents are filed both to strengthen and broaden the protection against competition (often called “fencing”) as well as to extend the effective duration of patent protection (Sternitzke 2013). Obtaining secondary patents is costly and can therefore be related to a company’s assessment of a project’s value. For this reason we expect the total number of patents surrounding a commercialization project to be positively related to the hazard of successfully reaching different stages in the process.

Finally, the number of claims a patent applicant includes in a patent application has been discussed as a value correlate as such claims mark the boundaries of a patent. However, it should be stressed that the arguments over how the number of claims is related to value are not uncontroversial: On the one hand, each additional claim might increase the probability of an infringement and, therefore, the value of a patent. On the other hand, however, additional claims make the description of the claimed invention more specific and might narrow the scope of the protected area and hence the value of the property right (Lanjouw and Schankerman 2004; Harhoff and Reitzig 2004). Without further assumptions it is therefore hard to predict how the number of claims in a patent affects the hazard of commercialization.

(b) Indicators related to the uncertainty surrounding patent protection

Patents are generally regarded as a highly effective means in appropriating the value from an invention in the pharmaceutical industry (Cohen, Nelson, and Walsh 2000). However, firms seeking to obtain a strong protection of their invention face several uncertainties. First, it is uncertain when and whether the applicant will obtain a patent at all. Hence firms are likely to condition their commercialization decisions on the strength of the patent protection

they can obtain. If, in the worst case, the applicant obtains no or only weak patent protection then commercialization will be unlikely. The timing of the actual patent grant should therefore play an important role for commercialization decisions. Only after a grant decision has been issued is uncertainty regarding the patentability and the scope of the granted patent resolved (Harhoff and Wagner 2009). In fact, Gans, Hsu, and Stern (2008) showed that the timing of the grant decision affects the economic outcome; in a study of out-licensing deals of technology start-ups they find that the hazard of licensing increases significantly after the grant decision has been issued. In the context of the commercialization of pharmaceutical inventions we expect similar effects: companies will be more willing to invest in commercialization activities once the risk of not obtaining patent protection (or of obtaining only a weak patent) as well as when the formulation of the claims included in the granted patent right has been resolved.

While the timing of the patent grant is a very direct measure of the uncertainty surrounding the patenting procedure the literature also identifies less direct indicators of uncertainty. In particular, critical references given by the European Patent Office (EPO) to prior art in its review of the patent application (Michel and Bettels 2001; Webb et al. 2005) have been used to characterize weaknesses in patents and patent portfolios. Critical references point to prior art that limits the patentability of an invention. An older, but similar, invention might threaten the patentability of a newer invention because the newer invention is not novel or lacks inventiveness. Critical documents containing conflicting prior art are then referenced by the EPO as X or Y references in a patent's search report. X refers to documents showing that a claimed invention cannot be considered novel or cannot be considered to involve an inventive step, while Y refers to the case where a document has to be combined

with one or more other documents to show lack of inventiveness.⁵ A series of papers have demonstrated that an increasing number of X and Y references not only decreases the likelihood of an application ultimately leading to a patent grant (Harhoff and Wagner 2009) but also increases the likelihood that a patent will be opposed conditional on the grant (Harhoff and Reitzig 2004; Harhoff, von Graevenitz, and Wagner, 2013). For this reason, it can be argued that X and Y references are indicators of legally “weak” patents as they reduce the likelihood of a patent grant but increase the likelihood of a post-grant validity challenge. A higher share of X or Y references among all references in a patent’s search report should therefore indicate a higher degree of uncertainty regarding the strength and/or the ability to obtain patent protection. On the other hand, it needs to be pointed out that X and Y-type references are an input to the patent approval process. During patent examination the patent applicant has the opportunity to revise the claims so that objections to patentability initially indicated by critical references are mitigated. Therefore X and Y references do not relate directly to the patent that is actually granted. Whether measures derived from critical references provide any information beyond whether the patent is likely to be granted and/or upheld after a patent opposition proceeding is currently an open question. We seek to shed light on this question by including the critical references into our empirical analyses.

⁵ See EPO Guidelines for Examination, Part B, Chapter X, Section 9.2.1, available at http://www.epo.org/law-practice/legal-texts/html/guidelines/e/b_x_9_2_1.htm, latest access on December 27, 2014.

3. Data and variables

The data used for our analysis are derived from two major sources: IMS Lifecycle R&D Focus database (“R&D Focus”) – as at April 2011 – and PATSTAT. R&D Focus is a proprietary database that contains extensive information on just under 30,000 development projects (i.e., product candidates) in the pharmaceutical and biotech industry until the end of 2009. It is organized by product, and each product record is compiled of information such as the developing company, the Anatomical Therapeutic Chemical (ATC) classification of the product, and a detailed history of the major clinical development and commercialization events.⁶ For about 30 percent of the products, R&D Focus lists the primary patent (or patents) covering the products. The primary patent covering a given product may be filed in any jurisdiction. In order to construct patent indicators that are comparable across products we use the patent family information provided by PATSTAT to identify the EP equivalent (if the primary patent is not filed at the EPO) of the primary patents listed in R&D Focus. Our sample includes all products contained in R&D Focus which are (1) either covered by a patent filed at the EPO or by a patent that has an EP equivalent filing, and (2) have priority dates greater or equal to 1980 for our study.

In total, we identify 5,923 products with an average of 1.12 primary patents related to them in R&D Focus (see Table 1). Since some of the non-EP primary patents have more than one EP equivalent filing, this results in a total of 8,247 unique EP patents with application dates ranging from 1980 to 2007. Moreover, some of the EP patents identified relate to more than one product, yielding a total of 9,229 unique product-EP patent-pairs. For each of the 5,923 products, we observe whether and when certain stages in the (pre-) clinical

⁶ Whether a product is included in the R&D Focus dataset prior to product launch depends to a large extent on voluntary disclosure by the product owner. Requirements for publicly listed companies to disclose “material” information often mandates disclosure of information about products in later-stage clinical trials, and in recent years new regulations have required more trials to be registered and disclosed (e.g., through clinicaltrials.gov). However, typically we only observe products that have reached a pre-clinical development milestone.

development process were achieved before the end of 2009. As Figure 1 shows, these products in our dataset are subject to the severe attrition along the various stages that is characteristic of drug development: Only 10.6 percent of the products are eventually launched in a European market, with an average duration from priority filing to market of slightly more than 11 years. These figures are broadly in line with earlier statistics derived from a comparable yet much smaller dataset in Sternitzke (2010).

INSERT FIGURE 1 ABOUT HERE

In our multivariate analyses we employ duration models that exploit the time between the invention of a product and reaching a particular stage in the development process as the dependent variable. This allows us to relate various indicators associated with an invention's value and the uncertainty surrounding patent protection to the hazard that a given stage of the development process will be reached, and thereby observe their effect on the duration of the development cycle. We obtain comprehensive data on the characteristics of the patents covering these products from PATSTAT, allowing use to compute these patent-based indicators. We focus on EP patents as there is more fine-grained information available than for patents filed at other patent offices. In cases where there is more than one primary patent associated with a project we use the primary patent with the earliest priority date for computing the patent characteristics. For each of these patents we compute the following indicators.

Legal status of the patent application

For each patent application we observe important features of its examination history, including the date when the application was filed at the EPO, the date it passed through subsequent stages in the patenting process, as well as when and if a patent grant (0/1) was issued (up to October 2011). In our multivariate analyses we employ duration models that

account for time-varying covariates. For this purpose we split each individual observation into several observations, dividing the data each quarter year (starting from the date of invention). For instance, an observation with an observed duration of 3.2 years would be split into 13 observations. We define a time-varying variable post-grant set equal to zero in the periods prior to the quarter in which a patent was granted and one in the period the primary patent was granted. The method is discussed in more detail below.

Characteristics of the patent document

We use the number of patents belonging to a patent family (family size) and whether the application was filed under the Patent Cooperation Treaty PCT (0/1) as indicators of the scope of international protection sought by the applicant. We include the number of claims listed on the patent as an indicator of the patent scope (van Zeebroeck, van Pottelsberghe de la Potterie, and Guellec 2009). Finally, we use the IPC classification (at the four-digit level) assigned by the patent office in order to capture structural difference across technologies.

Reference-based measures

Along with the publication of a patent application 18 months after its filing, the EPO also publishes the search report, which contains information on prior art relevant for the patentability of an application by referencing previous patents or non-patent literature (mostly scientific publications). We compute various indicators from these “backward“ references including the number of patent references and the number of non-patent references.⁷

In addition to the numerical count of patent references contained in the search report we also account for their composition. At the EPO, references contained in a patent’s search report

⁷ Various studies have also used the proportion of patent to non-patent references to characterize how close an invention is to basic science (Trajtenberg, Henderson, and Jaffe 1997; Narin and Noma 1985; Narin, Hamilton, and Olivastro 1997; Meyer 2000).

are classified into different categories: Most importantly, X-type references indicate that a single prior patent may undermine the novelty of the claims in the application. A Y-type reference may do the same in combination with other references. We compute the share of X-references and the share of Y-references (relative to the total number of references). Finally, we exploit the technology classification of the patent references to compute the originality index. The originality index is a Herfindahl-based index and is defined for patent i as

$$\text{originality}_i = 1 - \sum_k^{N_i} \left(\frac{\text{Nreferences}_{s_{ik}}}{\text{Nreferences}_{s_i}} \right)^2$$

where k is the index of technology classes and N_i is the number of different IPC4 classes to which the referred patents belong. Since $\text{Nreferences}_{s_{ik}}$ is the number of references contained in patent i that point to technology k and Nreferences_{s_i} is the total number of references, originality is bounded between 0 and 1. Higher values represent less concentration and hence more originality of the invention in the sense that it draws on a wider set of different technologies (Henderson, Jaffe, and Trajtenberg 1998).

Citation-based measures

The citations that a patent receive signal that the cited patent contributed to the state of the art in a technology field and hence the citation count is often used as a measure of the patent's value. For the patents in our sample we compute the number of forward citations these patents received from subsequent EP patents within a period of five years excluding self-citations.⁸ In addition, we also compute the number of self-citations a patent received within five years. This is the number of citations originating from subsequent patent filings by the same patent applicant. Finally, we also account for the distribution of the technology classification of the patents, citing a focal patent by computing the generality index analogous to the originality index. The generality of patent i is defined as

⁸ We compute the number of forward citation received within a 5-year period to avoid bias arising from the fact that "older" patents have a longer period over which they can receive citations compared to "younger" patents.

generality_i = $1 - \sum_k^{N_i} \left(\frac{N_{\text{citations}_{ik}}}{N_{\text{citations}_i}} \right)^2$, where k is the index of technology classes and N_i is the

number of different classes to which the citing patents belong. Generality is bounded between 0 and 1 with higher values indicating that a focal invention is relevant for a wider set of different technologies and is hence more general (Henderson, Jaffe, Trajtenberg 1998). It can be argued that inventions that are applicable to more technologies (and are thus more general) are more valuable than other inventions, therefore it is important to control for generality when including citation counts.

Applicant characteristics

Patent documents contain information on the patent applicant including its identity and its country of origin. The ECOOM-EUROSTAT–EPO PATSTAT Person Augmented Table (EEE-PPAT) provides a harmonization of the applicant names listed on EP patents.⁹ We use these harmonized names to compute the cumulative number of patent applications filed by a patent applicant on an annual basis as a coarse proxy of size. We also characterize applicants by their country of origin and distinguish between applicants located in Europe, in the US, in Japan or ‘the rest of the world’ (ROW). Finally, the EEE-PPAT data classify applicants according to their organizational form as companies, government/non-profit, university/hospital, and individuals. This information allows us to control for organizational form.

4. Descriptive statistics

Our sample consists of 8,247 unique EP patent filings corresponding to 5,923 products. In addition to figure 1 we report a breakdown of the different stages these products reached,

⁹ We use the applicant names provided by the ECOOM-EUROSTAT–EPO PATSTAT Person Augmented Table (EEE-PPAT) which provides harmonized applicant names for the PATSTAT database. See Magerman et al. (2010) for a full description.

distinguishing between entering pre-clinical trials, phase 1/2/3 clinical trials, and market launch in at least one major European country. Most notably, products which reached higher stages seem to be characterized by stronger patent protection than products which did not. Successful products are not only associated with a higher number of primary patents in R&D Focus, but also the EP equivalent applications corresponding to the R&D focus priorities are granted at a higher frequency. This finding might be driven by endogenous applicant behavior: Organizations will put more effort into the commercialization of valuable inventions, which might also drive their effort to obtain strong patent protection (Sternitzke 2010, 2013).

When interpreting Table 1, however, it must be taken into account that these findings might partly be influenced by censoring as products that reached advanced stages are older than the average product. As a consequence, the more advanced products are more likely to have completed the lengthy process of patent examination. This is one of the major reasons why we employ duration models in our multivariate analysis as they are able to properly account for this censoring of the data. It should be noted, however, that the number of priority patents related to a product is not affected by censoring but is a result of the organization's patenting strategy carried out in a relatively short time-period after invention. Taken together, Table 1 suggests that stronger patent protection is correlated with products that advance further in the development process.

INSERT TABLE 1 ABOUT HERE

It is important to note that only a small fraction of patent applications enter the pharmaceutical commercialization funnel. IMS R&D Focus lists about 30,000 pharmaceutical commercialization projects. At the same time a total of 398,957 patent applications have been filed in the same IPC4 classes as the 8,247 patents that are associated

with projects in the IMS database (we refer to these patents in the product sample). In Table 2 we present descriptive statistics on the product sample and also results from simple probit regressions in which we relate the likelihood that a patent from the underlying population is related to a project in the IMS database.¹⁰ Not surprisingly, we find that most value indicators are associated with a higher likelihood that the patent is included in the product sample. For instance, we find that both the number of jurisdictions where patent protection is sought after (family size) and the number of forward citations are positively related to a patent's inclusion in the product sample. Equally, we find that inventions that are more generally applicable (as indicated by the significant positive coefficient of generality) are more likely to be associated with commercialization efforts. Surprisingly, we also find that patents which are characterized by a higher share of type-X references are more frequently included in the product sample. Our interpretation of this counterintuitive finding is that type-X/Y references are not only a measure of uncertainty but are also a measure of patent breadth: The broader the patent claims are formulated in the application, the higher the number of type-X/Y references will be. Conditional on the grant, patents characterized by a large number of type-X references are therefore likely to be seeking a broader scope of protection, making it more attractive to commercialize the underlying inventions.

INSERT TABLE 2 ABOUT HERE

Finally, and most importantly, we relate patent characteristics that have been shown to impact outcomes within the patenting process to outcomes beyond the patent system. In Table 3 we report the descriptive statistics of the patents' indicators broken down by the stage that a product reached in the development process. Most value indicators correlate with a

¹⁰ Note that not all pharmaceutical projects in the R&D Focus database contain information on the underlying patent(s). In fact, only appr. 30% of all projects contain this information. For this reason, the control sample in the probit regression contains patents that are in fact related to projects.

product's commercialization success. In particular, the average family size and the number of forward citations are correlated with the product's stage in the development process. One notable exception, however, is the PCT filing path: The share of patents that have been filed under the PCT treaty decreases for products that are at higher stages in the development process. This finding is likely to be driven by truncation: The share of patents filed under the PCT regulation increases steadily over time (Harhoff and Wagner 2009). For this reason the share of PCT filing is higher in projects that enter the sample at a later stage. At the same time, projects entering the sample later are more likely to be censored. Hence, lower shares of PCT applications in projects reaching later stages of the commercialization funnel are the result of a truncation of the data. Regarding backward references, we observe that products that reach the later development stages are characterized by fewer backward references and a lower share of X references. We interpret this as an indication that firms are more likely to advance the development of products which are covered by patents at less risk of being invalidated by existing prior art in opposition proceedings or litigation in court.

INSERT TABLE 3 ABOUT HERE

Our descriptive statistics suggest there is a positive correlation between value indicators and successful product commercialization. At the same time, the descriptive statistics suggest that the strength of a patent is also related to successful commercialization. In the following, we analyze which of these patent-based indicators contain the most information with regard to the commercialization of products in a multivariate context. This allows us to draw conclusions as to what extent patent-based indicators are informative beyond the patenting process and are reliable proxies for outcomes in the product market.

5. Multivariate analyses

5.1 Methodology

In order to analyze the explanatory power of the patent-based indicators related to the value and uncertainty described above, we employ multivariate regressions to relate them to the hazard of reaching a particular stage in the product development process. In the absence of censoring, standard regression models such as Ordinary Least Squares (OLS) and probit would be suitable for modeling the effect of patent indicators on whether and when the underlying product reaches a particular stage. However, for a subset of the observations in our dataset we do not know the final outcome of the dependent variable. Our data on product commercialization only includes information until the end of 2009 and, given the long process of commercialization in the pharmaceutical industry, it is likely that a large share of the observations may still be under development. If standard regression models were employed, we would have to exclude these pending cases despite the fact that they convey information on process durations. Duration models, however, allow for this censoring by analyzing the marginal hazard at a point in time rather the average hazard across the whole period.

Unlike traditional regression models, duration models also allow us to incorporate the effect of time-varying variables. This is important as we are particularly interested in the effect of the exact timing of the grant of the primary patent on the speed of the project development. Once a patent has been granted, uncertainty surrounding the protective scope of the final patent is resolved and might therefore affect subsequent commercialization. In order to model the effect of the timing of the patent grant, we define the time-varying indicator post-grant that switches to one in the period a patent has been granted and all subsequent periods (see above). The variations in the duration until a primary patent is granted over different projects should allow us to identify the relation between the timing of a patent grant

on the speed of product commercialization. A similar approach was employed in Gans et al. (2008) to model the effect of a patent grant on the hazard of licensing taking place.¹¹

In our analysis we employ a Cox proportional hazards (PH) model with time-varying regressors.¹² The Cox PH model is specified as a semi-parametric model incorporating a non-parametric baseline hazard rate and a multiplicative term allowing regressors to have a proportional impact relative to the baseline (see Cox 1972; Kalbfleisch and Prentice 2002; or Kiefer 1988, for an overview of Cox PH models). Letting $\lambda(t, x)$ be the hazard of a certain event taking place in period t (i.e., the instantaneous probability of failure at t , conditional on survival until t) Cox PH models are generally specified as $\lambda(t, x) = \lambda_0(t) * \exp(x_t' \beta)$ with $\lambda_0(t)$ being the non-parametric baseline hazard rate, x_t the observed data (x_t indicates that the data matrix can include time-varying covariates) and β is a vector of unknown parameters (Cox 1972). Note that in all our tables we report the exponentiated coefficients $\exp(\beta)$, that is, the hazard ratios. The interpretation of hazard ratios is straightforward: The hazard ratio represents the proportional change in the likelihood of an event happening in t (conditional on it having not happened already) associated with a one unit increase in the underlying variable. Hence, hazard ratios larger than one indicate that a change increases the risk of an event taking place whereas hazard ratios less than one indicates a decrease in the risk (Kalbfleisch and Prentice 2002; Cleves et al. 2010).

We present a set of regressions specifying the time between invention and reaching a particular stage in the product development process as a function of the patent indicators described above employing Cox PH models (table 4). Here, we treat products that reached a

¹¹ If the timing of the patent grant was exogenous such a strategy would allow for the identification of causal effects (see Gans et al. 2008). However, it is a reasonable assumption that patent applicants can influence the timing of obtaining a patent grant by their level of cooperation with the patent examiner (see Harhoff & Wagner 2009). For this reason, we do not attempt to identify a causal effect here.

¹² In unreported robustness tests we also estimated Accelerated Failure time models and found similar results.

given stage in the commercialization process as completed spells. In particular, we run separate Cox PH regressions for the stages of preclinical trials, phase 3 clinical trials, and, ultimately, the market launch. Products that did not reach a given stage are included in these regressions but are treated as censored (Kalbfleisch and Prentice 2002; Cleves et al. 2010). The commercialization regressions include the time-varying post-grant indicator. The post-grant coefficient will allow us to identify the effect of the uncertainty being resolved. Finally, we also interact the post-grant indicator with the number of forward citations that the primary patent receives. The coefficient of this interaction term should be greater than one if uncertainty plays a more important role for more valuable projects.¹³

Note that in our regressions products which did not reach a given stage of the development process before the end of our observational period (December 2009) are treated as censored.¹⁴ All regressions include the fixed effects for the country of the type and origin of the patent applicant, the year of invention, the technology class (using the IPC level 4 classification) as well as the Anatomical Therapeutic Chemical (ATC) of the project. For reasons of brevity we do not report these results here.¹⁵

5.2 Results and discussion

Table 4 presents the results of the Cox PH regressions on the project level relating patent-based indicators to events beyond the patent system - that is, to the hazard that a project reaches a certain outcome in the product commercialization process. We run separate

¹³ We are grateful to the anonymous referee who suggested this approach.

¹⁴ For some observations we do not observe the dates of every stage. If we do not observe when a product entered a given stage we treat these observations as censored. In cases where we can infer that a product reached a given stage based on its observed entry into a subsequent stage, we treat them as censored and take the difference between the date they entered the next subsequent stage and their date of invention as their time at risk.

¹⁵ We observe a high variation in the fixed effects for ATC classes. For instance, in class B “blood and blood forming organs” (in particular B02 antihemorrhagics, B03 antianemic preparations) we find above average hazards of launching a product, whereas in classes J “antiinfectives for systemic use” and L “antineoplastic and immunomodulating agents” chances of successful commercialization are below average.

regressions for the hazard that a product has entered preclinical trials (columns 1 and 2), phase 3 clinical trials (columns 3 and 4), and whether it is ultimately launched in a European country (columns 5 and 6).

Overall, we find that patent-based indicators related to value and uncertainty are informative with regard to outcomes beyond the patent system and therefore allow us to draw inferences with regard to product market outcomes. Other patent characteristics, which we include as additional control variables in our regressions, add little explanatory power to the model and most of them are individually not significantly different from zero. Below we discuss the effect of the variables of main interest on the hazard that a particular stage is reached.

INSERT TABLE 5 ABOUT HERE

We formulated the expectation that inventions of higher value are correlated with higher commercialization hazards. By and large, our results confirm this expectation. Projects that are characterized by a higher number of patents, a larger family size, and a higher number of citations are also characterized by significantly higher hazards of eventually yielding a product that is launched on the market. One additional patent per product increases the hazard of reaching the market by more than 70 percent while an additional country in which patent protection is sought after (family size) is associated with a 5 percent higher hazard (see table 4, columns 5 and 6). The effects of the number of patents per product and family size on the hazard of reaching phase 3 clinical trials are of similar magnitude and are also highly significant (see table 4, columns 3 and 4). At the same time, we find that these two variables reduce the hazard of a product entering preclinical trials. The hazard ratio of an additional patent per product is smaller than one and translates into a 7.3 percent reduction of the hazard while an additional country is associated with a reduction of the hazard reaching preclinical

trials of -0.4 percent. We speculate that firms start preclinical trials only after developing a clear protection strategy, which is likely to take more time if more comprehensive protection (more patents per product and larger international scope) is being sought after. This might explain the differential effect of these variables being associated with a slower start of clinical trials but higher hazards of reaching the later stages. This explanation is also consistent with the uncertainty argument made in section 2. Finally, the number of claims regarding commercialization hazards is not informative. This is in line with prior literature arguing that the number of claims is hard to interpret and is not a direct measure of patent value (see above).

The results regarding citation-based measures of patent value point in a similar direction but are less clear. Patents characterized by a higher number of forward citations are related to a higher hazard of market launch. However, the effect is significant only at the 10 percent significance level (see table 4, column 5). The number of forward citations is not informative regarding earlier stages of the commercialization process. (Note, however, that the number of citations significantly moderates the effect of uncertainty, see below). Self-citations also have a measurable effect on commercialization outcomes. Patents with a higher number of self-citations are related to projects being significantly more likely to enter preclinical trials (see table 4, columns 1 and 2) but are also associated with lower hazards of reaching the market – though the latter effect is only marginally significant (see table 4, columns 5 and 6). To summarize, value indicators are related to higher hazards of successful commercialization. However, it becomes apparent that indicators that are based on a firm's patenting strategy, such as the number of patents or family size, are more informative than indicators based on citations that only arrive over time.

Our results also suggest that the uncertainty surrounding the patenting process has a significant effect on the hazards of product commercialization. Our main measure for

uncertainty relates to the timing of the grant of the primary patent associated with a project. The time-varying post grant indicator captures the timing of a patent grant (table 4, columns 1, 3, and 5). We additionally interacted the post grant indicator with the number of citations to scrutinize whether certainty plays a more important role for valuable inventions (table 4, columns 2, 4, and 6). The coefficient of the post grant indicator is positive and highly significant regarding the hazard of reaching preclinical and phase 3 clinical trials, highlighting that a resolution of uncertainty is associated with faster clinical testing. The effect on market launch is not significant, however. The interaction of the post-grant indicator with the number of forward citations reveals that a patent grant has a positive effect on the hazard of reaching preclinicals irrespective of a patent's value as the interaction term is insignificant in column 2 of table 4. For later stages, however, the resolution of uncertainty particularly effects valuable projects; the interaction effect is positive and marginally significant for phase 3 clinical trials and positive and highly significant for market launch (see table 4, columns 4 and 6). A potential explanation for this finding is that firms receive signals regarding a project's value only once it becomes clearer during clinical trials what medical indications can effectively be treated with a drug. Therefore, in early stages firms will have less precise information on a drug's potential value and the resolution of uncertainty should not have a differential effect on commercialization effort. At later stages, however, when more precise information on value becomes available firms seem to put more effort in the commercialization of valuable drugs once uncertainty regarding patent protection has been resolved.

The effect of critical references on commercialization outcomes is negligible. The share of type-X references is insignificant in all our regressions. The share of type-Y references is significant only in the regressions for phase 3 clinical trials. We have argued that these measures are based on the references contained in the search report underlying patent

examination. The final specification of the claims in a granted patent is likely to resolve some limitations to patentability indicated by critical references of the search report.¹⁶ Moreover, Harhoff and Wagner (2009) argue that critical references are highly significant predictors of whether a patent will be granted. As our regressions control for patent grant it is not surprising that these indicators have limited explanatory power in our regressions. It seems that critical references have limited predictive power beyond outcomes within the patent system.

6. Conclusion

Patent-based indicators are frequently used in empirical studies of innovation and technological change to describe not only phenomena related to the patent system itself but also to study outcomes from innovation, i.e., outcomes beyond the patent system. This paper contributes in various ways to our understanding of what can be concluded from patent-based indicators. A unique dataset allowed us to study how patent-based measures are related to different outcomes of product commercialization projects in the pharmaceutical industry. Analyzing the relation between patent-based measures and product commercialization shows that the relations are nuanced and need careful interpretation.

Our findings suggest that a number of patent-based indicators related to an invention's value and the uncertainty surrounding patent-based appropriation strategies are informative with regard to commercialization outcomes. First, family size and the number of patents surrounding a project (indicators generally considered to be positively associated with value) follow similar patterns: they are positively correlated with the speed of product launch, but negatively correlated with entering pre-clinical trials. At this stage, we can only speculate on what is driving the apparent contradiction. One potential explanation could be related to the

¹⁶ This explanation is backed up by informal interviews we conducted both with independent patent attorneys and IP professionals working for pharmaceutical companies.

uncertainty of the patenting process with regard to whether a patent will be granted and what its final specification will be, which is only resolved after a patent has been granted. For products with a higher value and/or a greater uncertainty of their final specifications, it is crucial to know the exact specification of the patent rights (and for the applicant to form expectations on how easy imitation by inventing around the patent will be) before it proceeds with clinical development. In these cases, the applicant might postpone decisions regarding product development until uncertainty has been resolved. Our empirical setting provides evidence for this explanation. The time-varying indicator, capturing the exact timing of when the primary patent associated with a commercialization project is granted, indicates that the resolution of uncertainty significantly speeds up the process. The effect is more pronounced for valuable projects as indicated by the interaction with the number of citations.

It has to be noted though, that these results have been obtained in the particular setting of the pharmaceutical industry. First, only a very small portion of all inventions (and hence patents) lead to actual commercialization projects and clinical trials. As we have shown, the patents underlying our product sample are a non-random sample from the population. Second, value appropriation in the pharmaceutical industry is based on a very small number of patents associated with final products. In other industries, a large number of patents (that are not always owned by one company) form the basis of a product. In these “complex” industries patent portfolios play a more important role than in the pharmaceutical industry (Harhoff, von Graevenitz and Wagner 2013).

In interpreting our findings, it is also important to note when and how the various measures are generated. In general, it is possible to construct some value-related measures such as family size, the number of claims, or whether the patent was filed through the PCT right after the patent application was published. Other value-related indicators, however, are generally not immediately observable. In particular, citation-based indicators can be

computed only several years after patent applications have been filed. Moreover, some of the measures are endogenous to private information the applicant has about the underlying product or what it learns during the patent prosecution process. If the patentee has private information regarding the probability that an invention will successfully pass through the clinical trials and/or has a high market potential he will be willing to invest more in obtaining a strong patent position. An applicant has 12 months to file international patent applications after the application (and 36 months on a patent filed through the PCT) and so may add (or remove) additional countries based on new information received regarding the likely success of the patent application or the underlying product. It may also adjust the number of claims during that time depending on the examiner's feedback given during the examination procedure. Meanwhile, the likelihood of being cited is likely to be affected by the product's success in preclinical and clinical trials, so it is not an exogenous indicator of success, although given that the average product enters preclinical trials only five years after priority filing, focusing on the first five years of citations limits the distortion.

Our study also has broader implications. It shows that uncertainty surrounding patenting procedures is associated with longer commercialization lags. As uncertainty seems to be particularly relevant for valuable inventions, our study underlines Régibeau and Rockett's (2010) argument that optimally a capacity constrained patent office should process patents related to important (and valuable) inventions faster than less important inventions. Budish, Roin, and Williams (2013) argue that fixed patent terms distort pharmaceutical companies' efforts in areas where clinical tests can be completed faster. In these areas companies can sell drugs for a longer period without threat of competition before patent protection expires. In light of Budish, Roin, and Williams (2013) one might fear that firms could be attracted to areas where patents can be obtained "relatively" easy independent of the social value of finding a cure for the underlying disease. Our study does not allow us to test whether such a

behavior is taking place or not. We hope that this question will be addressed in further research.

References

- Abrams, David S., Ufuk Akcigit, and Jillian Popadak (2013). Patent Value and Citations: Creative Destruction or Strategic Disruption? NBER Working Paper Series, no. 19647. Cambridge, MA.
- Adams, Christopher P. and Van V. Brantner. 2006. Estimating the Cost of New Drug Development: Is It Really 802 Million Dollars? *Health Affairs* 25(2), 420–428.
- Bessen, J. (2009). Estimates of patent rents from firm market Value. *Research Policy* 38: 1604–16.
- Budish, E., N. R. Benjamin, and H. Williams (2013). Do fixed patent terms distort innovation? Evidence from cancer clinical trials. NBER Working Paper Series., no. 19430, Cambridge, MA.
- Cleves, M., W. Gould, R. Gutierrez, and R. Marchenko (2010). An introduction to survival analysis using stata, Third Edition. Stata Press.
- Cockburn, I.M., and M. J. MacGarvie (2009). Patents, thickets and the financing of early stage firms: Evidence from the software industry. *Journal of Economics & Management Strategy* 18(3), 729–773.
- Cockburn, I. M., and M. J. MacGarvie (2011). Entry and patenting in the software industry. *Management Science* 57(5), 915–933.
- Cohen, W. M., R. R. Nelson, and J. P. Walsh (2000). Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not). NBER Working Paper, no. 7552. Cambridge, MA.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34(2), 187–220.
- DiMasi, J. A., R. W. Hansen, and H. G. Grabowski (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics* 22(2), 151–85.
- European Commission. (2009). Pharmaceutical sector inquiry – Final report. Brussels.
- Gambardella, A., D. Harhoff, and B. Verspagen (2008). The value of European patents. *European Management Review* 5, 69–84.
- Gans, J. S., D. H. Hsu, and S. Stern (2008). The impact of uncertain intellectual propertyrights on the market for ideas: Evidence from patent grant delays. *Management Science* 54 (5): 982–97.
- Gittelman, M. (2008a). A note on the value of patents as indicators of innovation: Implications for management research. *Academy of Management Perspectives* 22: 21–27.
- Gittelman, M. (2008b). Comment: The value of European patents. *European Management Review (D Journal)* 5: 85–89.

- Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28, 1661–1707.
- Hall, B. H., and D. Harhoff (2012). Recent research on the economics of patents. *Annual Review of Economics* 4, 541–565.
- Hall, B. H., A. Jaffe, and M. Trajtenberg (2005). Market value and patent citations. *RAND Journal of Economics* 36, 16–38.
- Harhoff, D., F. Narin, F.M. Scherer, and K. Vopel (1999). Citation frequency and the value of patented inventions. *Review of Economics and Statistics* 81, 511–515.
- Harhoff, D., and M. Reitzig (2004). Determinants of opposition against EPO patent grants – the case of biotechnology and pharmaceuticals. *International Journal of Industrial Organization* 22, 443–480.
- Harhoff, D., F. M. Scherer, and K. Vopel (2003). Citations, family size, opposition and the value of patent rights. *Research Policy* 32, 1343–1363.
- Harhoff, D., G. von Graevenitz, and S. Wagner (2013). Conflict resolution, public goods and patent thickets. *Management Science*, Forthcoming.
- Harhoff, D., and S. Wagner (2009). The duration of patent examination at the European patent office. *Management Science* 55, 1969–1984.
- Henderson, R., A. B. Jaffe, and M. Trajtenberg (1998). Universities as a source of commercial technology: A detailed analysis of university patenting, 1965–1988. *The Review of Economics and Statistics* 80, 119–127.
- Jaffe, A. B. (1986). Technological opportunity and spillovers of R & D: Evidence from firms' patents, profits, and market value. *The American Economic Review* 76, 984–1001.
- Kalbfleisch, J. D., and R. L. Prentice (2002). *The statistical analysis of failure time data* (Wiley Series in Probability and Statistics). Wiley-Interscience.
- Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature* 26(2), 646–679.
- Lanjouw, J. O., A. Pakes, and J. Putnam (1998). How to count patents and value intellectual property: The uses of patent renewal and application data. *Journal of Industrial Economics* 46, 405–432.
- Lanjouw, J. O., and M. Schankerman (2001). Characteristics of patent litigation: A window on competition. *RAND Journal of Economics* 32(1), 129–151.
- Lanjouw, J. O., and M. Schankerman (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *Economic Journal* 114, 441–465.
- Magerman T., J. Grouwels, X. Song, and B. Van Looy (2010). Data production methods for harmonized patent indicators: Patentee name harmonization. *EUROSTAT Working Paper and Studies*, Luxembourg.

- Malerba, F., and L. Orsenigo (1999). Technological entry, exit and survival: An empirical analysis of patent data. *Research Policy* 28(6), 643–660.
- Meyer, M. (2000). Patent citations in a novel field of technology — What can they tell about interactions between emerging communities of science and technology? *Scientometrics* 48, 151–178.
- Michel, J., and B. Bettels (2001). Patent citation analysis. A closer look at the basic input data from patent search reports. *Scientometrics* 51, 185–201.
- Moser, P. and J. Ohmstedt and P. Rhode (2013): Do Patent Citations Measure the Size of the Inventive Step? Evidence from Hybrid Corn. Available at SSRN: <http://ssrn.com/abstract=1888191>
- Mowery, D. C., J. E. Oxley, and B. Silverman (1996). Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal* 17, 77–91
- Narin, F., K. S. Hamilton, and D. Olivastro (1997). The increasing linkage between US technology and public science. *Research Policy* 26, 317–330.
- Narin, F., and E. Noma (1985). Is technology becoming science? *Scientometrics* 7, 369–381.
- Nerkar, A., and S. Shane (2003). When do start-ups that exploit patented academic knowledge survive? *International Journal of Industrial Organization* 21, 1391–1410.
- Pakes, A. (1986). Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica* 54, 755–784.
- Pakes, A., and M. Schankerman (1984). The rate of obsolescence of knowledge, research gestation lags, and the private rate of return to research resources. In *R&D, patents, and productivity*, ed. Z. Griliches. University of Chicago Press: Chicago.
- Putnam, J. (1996). *The value of international patent rights, economics*. Yale University, New Haven, CT.
- Régibeau, P., and K. Rockett (2010). Innovation cycles and learning at the patent office: Does the early patent get the delay? *The Journal of Industrial Economics* LVIII (2): 222–46.
- Sampat, B., and H. Williams (2014). How do patents affect follow-on innovation ? Evidence from the human genome. NBER Working Paper No. 19430.
- Schankerman, M., and A. Pakes (1986). Estimates of the value of patent rights in European countries during the post-1950 period. *Economic Journal* 96, 1052–1076.
- Sorenson, O., and T. E. Stuart (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology* 106(6), 1546–1588.
- Stuart, T. E. (1998). Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. *Administrative Science Quarterly* 43(3), 668–698.

Sternitzke, C. (2010). Knowledge sources, patent protection, and commercialization of pharmaceutical innovations. *Research Policy* 39, 810–821.

Sternitzke, C. (2013). An exploratory analysis of patent fencing in pharmaceuticals: The case of PDE5 inhibitors. *Research Policy*.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics* 21, 172.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology* 5, 19–50.

van Pottelsberghe de la Potterie, B., and N. B. Zeebroeck (2008). A brief history of space and time: the scope-year index as a patent value indicator based on families and renewals. *Scientometrics* 75(2), 319–338.

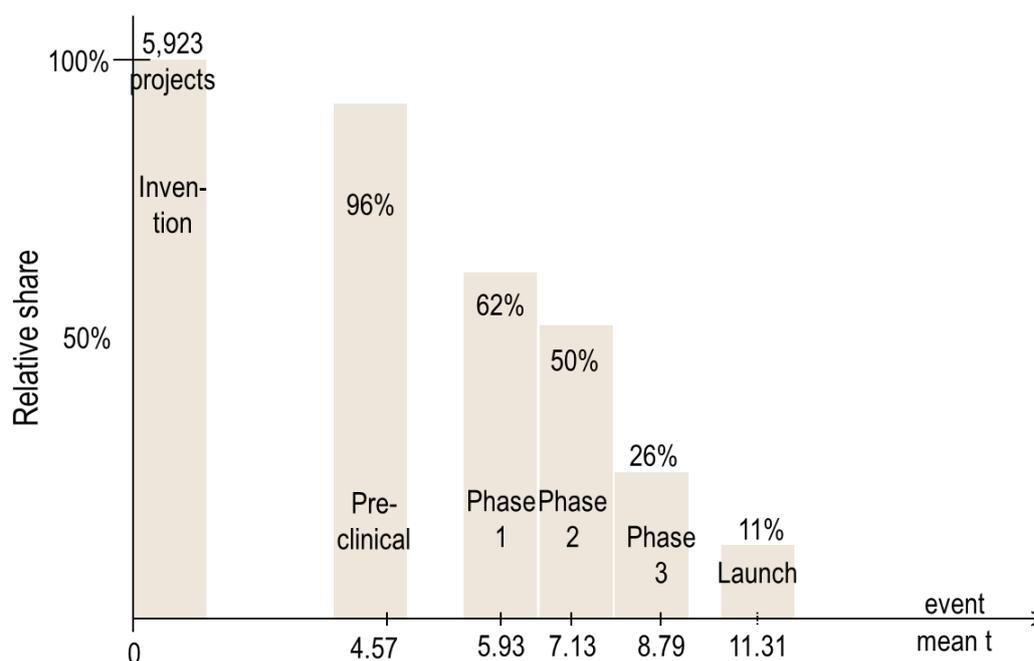
van Zeebroeck, N. B., B. van Pottelsberghe de la Potterie, and D. Guellec (2009). Claiming more: The increased voluminosity of patent applications and its determinants. *Research Policy* 38(6), 1006–1020.

Wagner, S., and I. Cockburn (2010). Patents and the survival of internet-related IPOs. *Research Policy* 39(2), 214–228.

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

Webb, C., H. Dernis, D. Harhoff, and K. Hoisl (2005). *Analysing European and international patent citations: A set of EPO patent database building blocks*. OECD Publishing.

Figure 1: Product attrition by stage of clinical development



Note: Clinical development of 5,923 pharmaceutical products for which we identified at least one European patent application. The vertical axis denotes the average share of products starting a given event during development. The horizontal axis denotes the average duration until a product reaches that stage.

Table 1: Number of products by stages in the product development process.

Development stage started	# Products	Reaching stage	Year of invention (estimated)	Age at reaching stage (years)	# Primary patents/product	# of EP patents	Granted	Opposed
Total	5,923	100%	1991.4	-	1.12	1.58	68.7%	7.05%
<u>Stage reached:</u>								
Preclinical	5,685	96.0%	1991.3	4.57	1.12	1.58	69.4%	6.83%
Phase 1	3,653	61.7%	1990.8	5.93	1.16	1.66	73.7%	7.93%
Phase 2	2,953	49.9%	1990.5	7.13	1.18	1.69	75.1%	8.50%
Phase 3	1,565	26.4%	1989.7	8.79	1.24	1.82	78.0%	11.43%
Launched in EP country	630	10.6%	1988.5	11.31	1.30	1.92	82.0%	17.66%

Note: The summary statistics regarding the patents related to the products are per product.

Table 2: Likelihood that a patent is associated with a project in the IMS database.

	Product sample (n=8,247)		Patent in product sample	
	(1) Means	(2) dx/dy (s.e.)	(3) dx/dy (s.e.)	
Patent characteristics				
Patent granted/upheld in opposition	0.69	0.0035** (0.0002)	0.0026** (0.0002)	
Filed through PCT	0.57	0.0011** (0.0002)	0.0006** (0.0002)	
Family size	15.58	0.0003** (0.0000)	0.0003** (0.0000)	
# Claims	19.79	0.0001** (0.0000)	0.0001** (0.0000)	
Citation-based measures				
# Forward citations (exc. self-citations) within 5 yrs (log)	1.87		0.0008** (0.0001)	
# Self-citation within 5 yrs (log)	1.00		0.0047** (0.0002)	
Generality index	0.24		0.0039** (0.0004)	
Reference-based measures				
# References to non-patent literature (log)	1.20	0.0009** (0.0001)	0.0008** (0.0001)	
# References (log)	3.21	-0.0002 (0.0001)	-0.0010** (0.0001)	
Share of X-type references	0.27	0.0015** (0.0003)	0.0011** (0.0003)	
Share of Y-type references	0.11	0.0002 (0.0004)	0.0003 (0.0003)	
Originality index	0.41	0.0004+ (0.0003)	0.0006* (0.0003)	
Applicant Characteristics				
Cumulative # applications (log)	15.33	-0.0010** (0.0000)	-0.0011** (0.0000)	
Applicant type (fixed effect)		YES	YES	
Applicant country (fixed effect)		YES	YES	
Fixed effects				
IPC4 dummy variables	-	YES	YES	
Year dummy variables	-	YES	YES	
N	-	398,957	398,957	
ll	-	-32343.8720	-31677.03	
chi2	-	15,614.89	16,962.12	
p	-	0.00	0.00	

Note: We report the average values for the 8,247 unique patents that are associated with projects in the IMS database. For the number of references, citations, and the cumulative number of patent applications we report mean values of the absolute values instead of the logs. Significance levels in the regression are denoted: + p<0.10, * p<0.05, ** p<0.01

Table 3: Patent-based indicators and outcomes beyond the patent system.

	# Products	Filed through PCT	Family size	# Forward citations (w/i 5 yrs)	# References to non-patent literature	# References to patents	Share of X-type references	Share of Y-type references
Total	5,923	56.3%	15.59	1.94	1.21	3.22	27.41%	11.07%
<u>Stage reached:</u>								
Preclinical	5,685	55.5%	15.78	1.93	1.19	3.21	26.81%	10.91%
Phase 1	3,653	53.0%	17.28	1.93	1.16	3.16	26.67%	11.06%
Phase 2	2,953	51.5%	17.07	1.97	1.15	3.14	26.17%	11.02%
Phase 3	1,565	48.2%	18.53	2.02	1.12	3.07	25.79%	11.81%
Launched in EP country	630	39.3%	21.00	2.33	1.10	2.86	22.53%	10.50%

Table 4: Cox Proportional Hazards regressions of the duration between filing and stages of the product development process on patent-based measures.

	Preclinical trials reached		Phase 3 clinical trials reached		Market launch in EP country	
	(1) HR/se	(2) HR/se	(3) HR/se	(4) HR/se	(5) HR/se	(6) HR/se
Legal status						
Post grant	1.254** (0.063)	1.296** (0.077)	1.208* (0.089)	1.088 (0.104)	1.165 (0.135)	0.927 (0.140)
Post grant * Number of citations		0.957 (0.040)		1.128+ (0.079)		1.288* (0.142)
Patent characteristics						
Number of patents/product	0.928* (0.035)	0.928* (0.035)	1.659** (0.092)	1.654** (0.091)	1.776** (0.140)	1.769** (0.139)
Filed through PCT	1.005 (0.041)	1.005 (0.041)	0.997 (0.087)	1.000 (0.087)	0.769+ (0.104)	0.773+ (0.104)
Family size	0.996** (0.001)	0.996** (0.001)	1.042** (0.003)	1.042** (0.003)	1.051** (0.004)	1.051** (0.004)
# Claims	1.000 (0.001)	1.000 (0.001)	0.999 (0.002)	0.999 (0.002)	1.001 (0.003)	1.001 (0.003)
Citation-based measures						
# Forward citations within 5 yrs (log)	1.008 (0.021)	1.018 (0.023)	1.022 (0.042)	0.965 (0.051)	1.111+ (0.068)	0.940 (0.090)
# Self-citation within 5 yrs (log)	1.066** (0.026)	1.066** (0.026)	1.004 (0.048)	1.007 (0.048)	0.872+ (0.063)	0.874+ (0.064)
Generality index	1.037 (0.072)	1.034 (0.072)	1.009 (0.148)	1.011 (0.148)	0.885 (0.197)	0.889 (0.199)
Reference-based measures						
# References to non-patent literature (log)	0.984 (0.023)	0.984 (0.023)	0.997 (0.048)	0.997 (0.048)	0.968 (0.071)	0.971 (0.071)
# References (log)	1.028 (0.024)	1.028 (0.024)	0.988 (0.048)	0.991 (0.048)	1.012 (0.077)	1.017 (0.078)
Share of X-type references	0.974 (0.043)	0.974 (0.043)	1.060 (0.099)	1.061 (0.099)	1.042 (0.153)	1.046 (0.154)
Share of Y-type references	0.930 (0.054)	0.929 (0.054)	1.626** (0.179)	1.628** (0.179)	1.303 (0.226)	1.317 (0.229)
Originality index	0.998 (0.042)	0.998 (0.042)	1.097 (0.097)	1.098 (0.097)	1.010 (0.139)	1.012 (0.139)
Applicant Characteristics						
Cumulative # applications (log)	1.006 (0.009)	1.006 (0.009)	0.973 (0.019)	0.972 (0.019)	1.049+ (0.030)	1.047 (0.030)
Applicant type (fixed effect)	YES	YES	YES	YES	YES	YES
Applicant country (fixed effect)	YES	YES	YES	YES	YES	YES
Fixed effects						
IPC4 dummy variables	YES	YES	YES	YES	YES	YES
ATC dummy variables	YES	YES	YES	YES	YES	YES
Year dummy variables	YES	YES	YES	YES	YES	YES
Observations	164,641	164,641	387,753	387,753	438,456	438,456
Log-likelihood	-42,755.4	-42,754.8	-10,727.2	-10,725.7	-4,457.8	-4,455.1
chi2	1,144.11	1145.23	840.12	843.12	806.95	812.29
P-Value	0.00	0.00	0.00	0.00	0.00	0.00

+ p<0.10, * p<0.05, ** p<0.01

Note: We report exponentiated coefficients, i.e., hazard ratio HR, in the table. Significance levels testing whether the underlying coefficients are equal in the table are denoted: + p<0.10, * p<0.05, ** p<0.01.