

Determinants of pharmaceutical innovation: the role of technological opportunities revisited

Bastian Rake¹ 

Published online: 5 August 2017
© Springer-Verlag GmbH Germany 2017

Abstract Recent empirical contributions emphasize the importance of (potential) market size for the development of new pharmaceuticals. At the same time, many scholars point out the importance of scientific advances for the industry’s R&D activities. Against this background I analyze the relationship between (potential) market size, technological opportunities, and the number of new pharmaceuticals in the United States. Technological opportunities are operationalized as growth rates of the relevant knowledge stock, as proposed by Andersen (Struct Chang Econ Dyn 9(1):5–34 1998, J Evol Econ 9(4):487–526 1999). I analyze a unique dataset by using an “entry stock” Poisson quasi-maximum likelihood estimator. The results reveal a rather robust and significantly positive response of the number of new pharmaceuticals, i.e., new molecular entities or new drug approvals, to market size and technological opportunities.

Keywords Determinants of innovation · Pharmaceuticals · Demand · Technological opportunities

JEL Classification O31 · J10 · J20 · L65

1 Introduction

There has been an extensive and long-running discussion as to whether demand- or supply-side factors determine innovative activities across different industries. The

✉ Bastian Rake
bastian.rake@uni-passau.de

¹ Chair of Organization, Technology Management and Entrepreneurship (OTE),
University of Passau, 94030 Passau, Germany

controversy between advocates of the “demand-pull” and those of the “technology-push” theories has been resolved by pointing out the importance of both the information flows of scientific and technical knowledge and information from users of products and processes for innovative activities (Freeman 1994). Consequently, this paper studies the contribution of demand-side variables (market size) and supply-side variables (technological opportunities) to innovative activities using data from the U.S. pharmaceutical industry. On the one hand, the importance of demand for innovation in this sector has been pointed out by many scholars (e.g., Cerda 2007; Acemoglu and Linn 2004). On the other hand, there is much evidence that innovative activities in this sector rely on scientific advances, particularly in the field of biotechnology (e.g., Rasmussen 2010; McKelvey 1996; Gambardella 1995). Hence, the pharmaceutical industry provides an almost ideal case to analyze the importance of demand- and supply-side variables for innovative output.

While researchers studying the determinants of innovation can apply a broad range of usually quite well defined demand-side variables, some important supply-side variables lack a clear definition or operationalization. In particular, the concept of technological opportunities has been found to be a key variable for explaining innovative activities across different industries, but there is no clear consensus concerning its empirical approximation (e.g., Klevorick et al. 1995). Early studies of the importance of technological opportunities relied mostly on industry dummies or survey-based indicators of the importance of different knowledge sources. In contrast to these contributions, I will use a quantitative proxy for technological opportunities, following an idea proposed by Andersen (1998, 1999), and operationalize technological opportunities by the growth rate of the knowledge stock in a given field of research.

The empirical analysis is performed on a unique dataset that encompasses data from different sources. Data concerning new pharmaceuticals come from the Drugs@FDA database. Clinical trials data are obtained from ClinicalTrials.gov. Market size is constructed using the March Supplement of the U.S. Current Population Survey and the Medical Expenditure Panel Survey (MEPS).¹ Publication data from the Web of Science databases is used to construct proxies for technological opportunities. The data are analyzed using an “entry stock” Poisson quasi-maximum likelihood estimator (QMLE), as suggested by Blundell et al. (1995), since it accounts for dynamic feedback and unobserved heterogeneity. Moreover, the Poisson QMLE is consistent under the assumption of a correctly specified mean and does not require Poisson-distributed data.

The results indicate a significantly positive response of the number of new pharmaceuticals to market size. The positive response holds true for new molecular entities as well as for new drug approvals and is fairly robust against different specifications

¹This paper predominantly uses data from the U.S., one of the largest markets for pharmaceuticals in the world. Many new pharmaceuticals introduced in the U.S. are marketed worldwide and approval decisions in many countries are based on the corresponding decision by U.S. authorities (Kremer 2002). Nevertheless, one should be careful with drawing conclusions for the relationships in the world market for pharmaceuticals.

of the market size proxy. Technological opportunities have a significantly positive relation with the number of new pharmaceuticals. However, the analysis does not reveal a robust significant relation between technological opportunities in biotechnology and basic research and the number of new pharmaceuticals. This result may be driven by the construction of the database and the knowledge stocks, which may better account for applied research. I take time lags of seven years into account, which corresponds to the average time needed from the first clinical testing of a promising compound to the approval of a new drug. The results suggest no clear-cut relationship between lagged market size, lagged technological opportunities, and the number of pharmaceutical innovations. By instrumenting market size by market size five years previously, I find no evidence for reverse causality. The number of phase II and phase III clinical trials, a proxy for new compounds under development, respond positively to (potential) market size and to technological opportunities at the beginning of the clinical research process.

This paper contributes to the literature in three ways. First, the relationship between demand and pharmaceutical innovation is often analyzed on a rather aggregated level, e.g., drug categories or disease groups. Instead, I use a more disaggregated level of analysis, namely medical indications. Second, by operationalizing technological opportunities as growth rates of knowledge stocks, I add a quantifiable proxy for this variable to the analysis. Third, the publication data used to construct the knowledge stocks are more closely related to the definition of technological opportunities than are patents or other types of data previously used in the literature. Particularly in a science-based industry such as pharmaceuticals, publications may reflect the possibilities for innovative activities more directly than, e.g., patents, which can be seen as a realization of specific possibilities for innovation. This interpretation is supported by a high number of citations from patents to bio-pharmaceutical publications (cf., Narin and Olivastro 1992) indicating that knowledge in scientific publications in this field of research may, on average, be more basic, occur prior to patented knowledge, and may contain knowledge that cannot be patented. Additionally, the number of approved drugs and the number of clinical trials provide a more appropriate measure for the industry's innovative activities as compared to patent counts. Due to the specificities of the industry's R&D process, almost all promising compounds are patented before their therapeutic potential is fully examined, but only few new compounds are approved for marketing, as the vast majority usually do not meet expectations in terms of safety and efficacy. Hence, the number of patents might considerably overestimate the number of successfully developed products. As a potential proxy for scientific and technological development, patents face the problem that classification schemes are usually oriented towards the chemical structure of compounds and not towards disease specific applications. Moreover, the specific therapeutic use may not be fully determined when the patent application is filed.

This paper is structured as follows. Section 2 reviews the related literature on the demand- and supply-side determinants of innovative activities. Section 3 presents the empirical strategy and Section 4 explains the construction of the variables and the different data sources used. The empirical results can be found in Section 5. Finally, Section 6 concludes.

2 Related literature

2.1 Demand and innovation

The importance of demand in explaining the rate of innovation across sectors has been recognized in economics for many decades. In his seminal contribution, Schmookler (1966) suggests that cycles of capital good patents usually lag behind cycles of capital goods' output and capital expenditures in downstream industries. This finding led to the conclusion that demand side stimuli are the most important for explaining the rate and direction of inventive activities. Along with other early contributions, these results have been criticized in the "demand-pull" versus "technology-push" debate. The corresponding studies often lack a clear definition of demand, neglect supply-side influences, and do not conclusively demonstrate the importance of demand for the rate and direction of innovation (Mowery and Rosenberg 1979). Replications of Schmookler's (1966) results have pointed to a somewhat weaker relationship between demand and innovative activities than in the original study. However, the relation remains significant in most cases (Scherer 1982). Furthermore, there may be a mutual dependency between demand and innovation, rather than a clear, unidirectional causality (Kleinknecht and Verspagen 1990).

Analyzing innovative activities in the United Kingdom, Geroski and Walters (1995) find that variations in demand Granger-cause major innovation counts and patents. Walsh (1984) offers a more differentiated view, by analyzing sub-sectors of the chemical industry in more detail. The results suggest that exogenous major innovations create their own demand, which subsequently leads to a rising number of incremental innovations. Thus, the importance of supply- and demand-side influences may vary over a life cycle. In line with these results, Barge-Gil and López (2014) report that demand plays a more important role in shaping companies' development rather than their research activities. In some industries, heterogeneous demand as well as lead or experimental users play a dominant role in shaping the rate and direction of technological development (Malerba et al. 2007; Adner and Levinthal 2001; von Hippel 1976).

As to the pharmaceutical industry, it is often argued that demand-side variables determine the rate and the direction of the industry's investment in R&D activities and, consequently, the number of new drugs in different therapeutic categories. Empirical evidence suggests that pharmaceutical companies' R&D decisions are determined by the expected return and cash flow variables, which should be connected with the size of the corresponding markets and with real drug prices (Giacotto et al. 2005; Grabowski and Vernon 2000). Expected returns are an important determinant of R&D spending not only for large U.S. but also for Japanese pharmaceutical companies (Mahlich and Roediger-Schluga 2006). These results support earlier findings, indicating that variables related to the expected returns explain chemical and pharmaceutical firms' R&D intensity (Grabowski 1968).

Following the empirical evidence on variables affected by market size shaping at least partly firms' R&D decisions, one would expect firms to invest in R&D and to introduce new pharmaceuticals, particularly for medical conditions that have a high expected market size and consequently, high expected returns. Consistent with

this argumentation Fabrizio and Thomas (2012) provide evidence that pharmaceutical companies' innovation activities respond not only to worldwide sales but also to those in their home countries. Acemoglu and Linn (2004) analyze the effect of (potential) market size on new drugs approved by the U.S. Food and Drug Administration (FDA), making use of exogenous changes in the potential market size driven by demographic trends. Their results indicate that a one percent increase in the potential market size of a drug category is associated with a four to six percent increase in the entry of new molecular entities. The response as to new non-generic drugs is approximately four percent. Moreover, their results show responses to five-year leads in market size. Taking the time needed for development into account, pharmaceutical R&D seems to respond to anticipated changes in the market size with a lead of 10 to 20 years. The results are robust against the introduction of controls for a variety of supply-side factors.

Using U.S. demographical data, Cerda (2007) provides additional evidence for the importance of market size for the introduction into the market of new molecular entities. Moreover, the author shows that there are important feedback effects, since new drugs affect the market size through their impact on the mortality rate. Market size related variables affect not only the introduction of new molecular entities and new non-generics, but also generic entry (Reiffen and Ward 2005; Morton 1999).

The importance of market size for the rate and direction of innovative activities related to new drugs has been particularly emphasized regarding diseases that affect individuals in developing countries. Small markets make it less attractive for pharmaceutical companies to engage in research and drug development for the corresponding medical conditions and the appropriate delivery of medications in developing countries (Lichtenberg 2005a; Kremer 2002). Similar problems may occur regarding rather rare diseases in developed countries. Supply-side incentives such as tax credits may lead to an increase in innovative activities related to such medical conditions. However, firms may still concentrate on the more prevalent conditions among the rare diseases (Yin 2008).

2.2 Technological opportunities and innovation

The importance of demand-side variables has been contradicted by advocates of the technology-push hypothesis, claiming that scientific and technological progress are the primary drivers of the rate and direction of innovative activities. The generation of new scientific and technological knowledge leads to a steady renewal of the pool of technological opportunities, i.e., an industry's set of possibilities for innovative activities (Klevorick et al. 1995). Technological opportunities reflect the state of knowledge at a particular point in time, which determines the cost and difficulty of successful innovative activities (cf., Jaffe 1986). Technological opportunities are bound to technological paradigms determining the scope of potential innovations and the effort necessary to achieve these along specific trajectories. New paradigms generate new opportunities for previously infeasible product development and productivity increases. It is, however, the expected economic returns that lead to dedicating resources to the exploitation of the existing opportunities (Dosi 1988).

The theoretical contributions clearly indicate that technological opportunities are a key factor in explaining (persistent) differences in innovative activities across industries (e.g., Nelson 1988). However, empirical investigations of this widely accepted view turned out to be more difficult to conduct since there is no clear consensus concerning the precise operationalization of the concept of technological opportunities (Cohen and Levin 1989). Consequently, a broad variety of proxies for technological opportunities has been used in empirical research. Many studies rely on industry categories or technology groups using different levels of aggregation. Although these binary variables may capture more than mere differences in technological opportunities, they explain a considerable fraction of inter-industry and inter-firm differences in innovative activities (e.g., Jaffe 1986, 1989a; Scherer 1965).

Other studies have relied on survey data addressing the sources of knowledge at the firm or industry level, particularly the importance of various basic and applied sciences for the industries' technological change, as well as the contribution of external knowledge sources, such as universities, suppliers, and customers (Barge-Gil and López 2014; Klevorick et al. 1995; Levin et al. 1987). These studies have found that technological opportunities, particularly the closeness to science and extra-industry sources of knowledge, positively affect innovative - particularly research as opposed to development - activities and the rate of technical change (Barge-Gil and López 2014; Nelson and Wolff 1997; Cohen and Levinthal 1989; Cohen et al. 1987; Levin et al. 1985). Another approach has used patent data to analyze technological opportunities and their development. Andersen (1998, 1999) defines technological opportunities as the growth of the stock of patents in different technology groups. The analysis of technology dynamics reveals opportunity differences across technologies and their changing sources over time.

The pharmaceutical industry is traditionally regarded as an industry with a rather high level of technological opportunities. Particularly, the transformation from "random screening" to a "guided drug discovery" process, and advances in molecular biology and genetic engineering, have opened up new technological opportunities (Galambos and Sturchio 1998; McKelvey 1996; Gambardella 1995). A broad variety of studies have illustrated the importance of scientific research, particularly that conducted in universities and public research institutions, for the creation of new knowledge. This new knowledge leads to the renewal and expansion of the available set of technological opportunities. The set of technological opportunities itself builds a foundation for the development of new pharmaceuticals (e.g., Cohen et al. 2002; Klevorick et al. 1995; Mansfield 1991; Jaffe 1989b). In line with these results, public research funding has been found to be another supply-side variable influencing the introduction of new pharmaceuticals to the market, since it should translate into the generation of new knowledge (Toole 2012).

Against this background, it seems rather surprising that supply-side variables, and among them particularly the concept of technological opportunities, have received relatively little attention from scholars analyzing innovative activities in the pharmaceutical industry. Griliches et al. (1991) show that technological opportunities linked to patent activity are positively related to the pharmaceutical industry's innovativeness. In order to control for different technological opportunities when analyzing the influence of organizational competence on firms' research productivity,

Henderson and Cockburn (1994) use dummy variables for therapeutic classes. While clearly focusing on the (potential) market size as the main determinant of pharmaceutical innovation, Acemoglu and Linn (2004) use pre-existing trends in the number of drug approvals to control for scientific opportunities across therapeutic categories.

Despite numerous theoretical arguments emphasizing the importance of technological opportunities for pharmaceutical innovation, difficulties with the operationalization of this concept may be one reason for the marked scarcity of empirical evidence in the literature. By using growth rates of knowledge stocks as a proxy for technological opportunities, as proposed in Andersen (1998, 1999), I will address this shortcoming of the previous empirical literature.

3 Empirical strategy

The empirical analysis draws upon the literature concerning the “demand-pull” versus “technology-push” debate, and takes into account demand- and supply-side factors as the explanatory variables for pharmaceutical innovation. This approach goes beyond recent contributions to the field, which mainly concentrate on the demand-side effects (e.g., Cerda 2007; Acemoglu and Linn 2004). Innovation is measured as the number of new drugs approved by the FDA over time.² Hence, a Poisson model for the conditional mean of new pharmaceuticals can be derived:

$$E[N_{it}|\zeta_i, X_{it}] = \exp(\beta_0 + \beta_1 \ln(MK S_{i,t-a}) + \beta_2 T O_{i,t-b} + \beta C_{i,t-c} + \zeta_i + \mu_t). \quad (1)$$

Here, N_{it} denotes the number of new pharmaceuticals for medical indication i in year t , ζ_i represents time-constant medical indication intercepts capturing heterogeneity among diseases, and X_{it} refers to all explanatory variables at t and earlier. More precisely, $MK S_{i,t-a}$ accounts for the (potential) market size in i at $t - a$, $T O_{i,t-b}$ refers to the corresponding technological opportunities, $C_{i,t-c}$ is a vector of control variables, and μ_t is a set of dummy variables for each year included in the analysis to capture common year specific effects.

Pharmaceutical innovation is a dynamic and non-linear process. Econometric models used to analyze innovative activities in the drug industry have to take this into account. More precisely, dynamic feedback can occur, since scientific breakthroughs in t may influence technological opportunities in t and subsequent years and, hence, innovative output. Given the typical time structure of the industries' R&D and innovative activities, possible effects on innovation may occur around 12 years after the original discovery, or even later. The presence of feedback processes violates the strict exogeneity assumption of conventional fixed or random effects panel

²The definition of innovation as a newly approved drug uses the specificities of the innovation process in the pharmaceutical industry. After several phases of pre-clinical and clinical testing, a new drug has to be approved by a regulatory body, the FDA in the United States, before being marketed (see, Gambardella (Gambardella 1995)). This allows an identification of the number of new drugs approved for marketing in a specific year. Additionally, the number of clinical trials in a specific phase can be used in a robustness check as a proxy for the number of new compounds under development.

estimators. Therefore, I follow Toole (2012) in applying an “entry stock” Poisson QMLE, as suggested by Blundell et al. (1995), which accounts for feedback and unobserved heterogeneity among medical indications. Monte Carlo simulations show that this empirical approach performs quite well if the sample size is rather small and the explanatory variables are highly persistent (Blundell et al. 2002). A high level of persistence over time can be found particularly for the market size measures. According to this approach, Eq. 1 can be transformed to

$$E [N_{it} | \bar{N}_{ip}, M_{ip}, \bar{T}O_{ip}, X_{it}] = \exp(\beta_0 + \beta_1 \ln(MKS_{it-a}) + \beta_2 T O_{i,t-b} + \beta C_{i,t-c} + \phi \bar{N}_{ip} + \theta M_{ip} + \rho \bar{T}O_{ip} + \mu_t), \quad (2)$$

Different pre-sample variables are included in the analysis in order to consider different sources of heterogeneity among medical indications such as the ease of introducing drugs addressing a particular medical indication or the level of competition in the market as expressed by the number of competing drugs (Toole 2012). More precisely, the pre-sample mean of new pharmaceuticals \bar{N}_{ip} has been included to account for differences in innovative activities. Differences in the burden of disease and in the technological opportunities may be additional sources of heterogeneity across medical indications. These aspects are taken into account by including the pre-sample prevalence of each medical indication i (M_{ip}) as well as their pre-sample technological opportunities ($\bar{T}O_{ip}$) into the analysis.

The Poisson QMLE is applied to a population average model in order to estimate (2).³ The population average model accounts for correlations over time for the medical indications included in the analysis (cf., Cameron and Trivedi 2013). The Poisson QMLE is consistent under the assumption of the correct specification of the conditional mean. This implies that the data need not be Poisson distributed (cf., Gourieroux et al. 1984). Throughout the analysis, I use Huber–White robust standard errors in order to account for the possibility of overdispersion and non Poisson-distributed data.

4 Data and variable construction

4.1 Pharmaceutical innovation

Data on new drugs were obtained from the U.S. Food and Drug Administration (FDA) using the Drugs@FDA database.⁴ Each record contains information on the trade name of the drug, its generic name, the components that provide its pharmacological activity, its approval date, and a classification of the newness of the drug formulation. I construct two different proxies for the number of innovations N_{it} . The

³This specification does not allow computation of traditional goodness of fit measures such as the Akaike Information Criterion. The quasi-likelihood under the independence model criterion (QIC) provides an alternative for Poisson quasi-maximum likelihood models (Pan 2001). Since it is not available for all model specifications used throughout the analysis, the results are available on request.

⁴<http://www.fda.gov/Drugs/InformationOnDrugs/ucm135821.htm>.

first follows most of the literature on pharmaceutical innovation, by defining an innovation as a new molecular entity (NME) (Cerdea 2007; Grabowski and Wang 2006). This measure refers to an active ingredient that has never been marketed before in the United States. Second, I use the number of new drug approvals (NDA) as a broader measure of pharmaceutical innovation. This proxy encompasses NMEs as well as new chemicals derived from marketed active ingredients, new formulations and dosages of already marketed ingredients, and new combinations of marketed ingredients in the year of their approval by the FDA.⁵

The data on new drugs were matched with the diseases or medical indications used in this paper based on their active ingredients which determine their therapeutic use. Therefore, I follow Cerda (2007) in consulting the 19th edition of the Drug Information Handbook published by Lexi-Comp and the American Pharmaceutical Association (Lacy et al. 2010). This handbook is comparable to a pharmaceutical dictionary, providing a list of drugs' active ingredients, the medical conditions for which the drug is used, and further information such as adverse effects. I take into account only those medical conditions that can be found on the FDA approved label. Hence, unlabeled and investigational uses are not taken into account. For the period 1974 to 2008, 530 unique NMEs and 1,484 unique NDAs had been approved by the FDA in the 168 diseases or medical indications analyzed in this study. However, an NME or NDA may be used as therapy for several medical indications. In this case, an NME or NDA is counted as an innovation for all the medical indications for which it is approved. Figure 1 shows the number of unique NME and NDA throughout the years included in the analysis.⁶

4.2 Clinical trials

New pharmaceuticals have to go through several phases of clinical testing before they can be approved for marketing by the regulatory authorities. I make use of this specificity of the pharmaceutical innovation process and use the number of phase II and phase III clinical trials as an alternative indicator for innovative activities in a robustness check. If market size and technological opportunities are related to the number of new drugs, they should also be related to the number of clinical trials used to test drug candidates. Although a single drug candidate is usually tested in several clinical trials, the number of clinical trials might serve as a proxy of the number of new drug candidates that might be approved for marketing in the future. Phase II ($Ph2Trials_{it}$) and phase III ($Ph3Trials_{it}$) focus mainly on the effectiveness of new compounds involving an increasingly larger number of more diverse test persons over longer periods of time. During these phases, accurate profiles of the tested

⁵Kneller (2010) shows that that new drugs originate predominately in pharmaceutical and biotechnology companies. Notwithstanding the important role of knowledge generated in universities and public research institutes for successful innovative activities in the pharmaceutical industry in general, academic institutions play only a minor role as originators of new pharmaceuticals. Applications for market approval are usually filed by pharmaceutical and biotechnology companies, too.

⁶Table 7 in the Appendix provides more descriptive statistics for the variables used in the analysis.

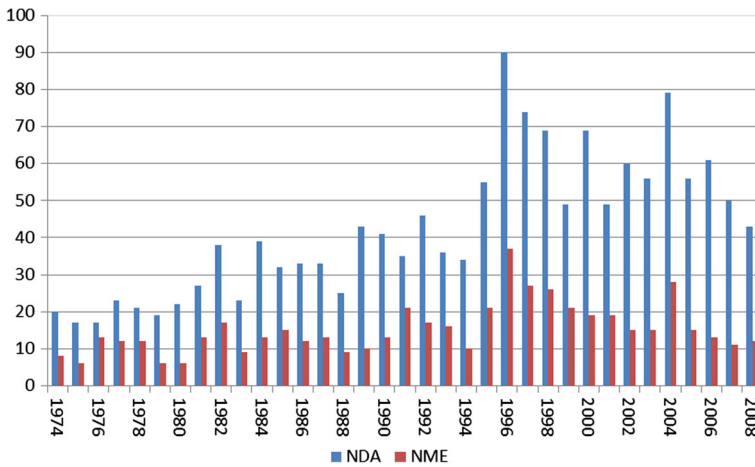


Fig. 1 Number of NDA and NME per Year, Source: Drugs@FDA

drug candidates are developed, including dosages and adverse effects occurring with small probabilities or after longer time periods. For a more detailed overview see, Gambardella (1995).

The clinical trials data are collected from ClinicalTrials.gov, a comprehensive registry of clinical trials maintained by the United States National Library of Medicine at the National Institutes of Health. It was established in 2000 and contains detailed information on each clinical study including design of the study, the locations where it is conducted, and the disease and condition that is addressed. Clinical trials listed in the database are conducted in the U.S. and over 180 other countries. The conditions specified in the clinical trials data have been used to match clinical trials with the medical indications used in this paper. Concentrating on drug related clinical trials that have been started during the years 2000 to 2008 I obtain 6,160 phase II and 4,622 phase III unique clinical trials.

4.3 Market size

Similar to related studies, market size is constructed based on demographic data using population sizes in different age groups and individual income. In doing so, I link market size to the potential purchasing power of individuals since a higher income allows higher drug expenditures (cf., Cerda 2007; Acemoglu and Linn 2004). There are two arguments for the use of demographic data in the construction of the market size measure. First, drug consumption is likely to change with age. The elderly may spend more on drugs than do younger individuals. However, it is very likely that they are affected by different medical indications than younger members of the population. Second, market size variations due to demographic changes are largely exogenous to scientific achievements, the entry of new drugs, and the endogeneity of market size and innovation (Acemoglu and Linn 2004). Therefore, market size per medical indication is proxied by the sum of the purchasing power of five-year

age groups a from 0–4 to 85+ for different medical indications, using appropriate weights. More precisely, market size per medical indication i in year t is defined by

$$MK S_{it} = \sum_a \omega_{ia} pop_{at} inc_{at}. \quad (3)$$

Here, ω_{ia} is a weight for medical indication i in age group a , pop_{at} is the total population in age group a at t , and inc_{at} is the individual income of persons in age group a in year t .

Following Acemoglu and Linn (2004), I compute ω_{ia} as the average share of expenditure for drugs assigned to indication i divided by the total individual income of individuals in age group a . To construct this proxy, I use data on drug expenditures and income from the MEPS for the years 1996 to 1998.⁷ I use the 9th revision of the International Classification of Diseases Clinical Modification (ICD-9-CM) published by the National Center of Health Statistics to match the data on drug spending to the medical indications included in this study. The weights are not time varying, in order to rule out changes caused by new possibilities of diagnosing a disease, new pharmaceuticals with which to treat it, and other causes of differences in the prevalence rates.

Population sizes per age group and year (pop_{at}) were obtained from the March Supplement of the U.S. Current Population Survey (CPS) provided by the Bureau of the Census for the Bureau of Labor Statistics. The CPS monthly surveys the members of about 50,000 households selected to represent the U.S. non-institutional population. Each member of the surveyed households is assigned to its age group according to the reported age. In addition to the demographic variables, the March Supplement provides information about income. Individual income inc_{at} is proxied by the median of the household income divided by the number of persons in the household. I use this proxy instead of the reported individual income because the latter hardly allows computing meaningful market size measures for those groups that have no or only very little own-income, particularly children, students, and non-working spouses. The income data are adjusted for inflation using the U.S. Consumer Price Index (CPI) for all urban consumers provided by the Bureau of Labor Statistics. The logarithms of the market size measures are what enter into the regression analysis.

4.4 Technological opportunities

It is widely accepted in the literature on technological change that technological opportunities play an important role in explaining differences in innovative activities across sectors and over time. However, there is no general consensus concerning the precise definition of this theoretical concept, and particularly its empirical operationalization. Not surprisingly, empirical research has applied a broad variety of indicators for technological opportunities, e.g., industry dummies, survey based measures on the sources of technical advance, and innovation expenditures as share of total expenditures (e.g., Castellacci 2007; Becker and Peters 2000; Nelson and Wolff

⁷This is exactly the timespan used in Acemoglu and Linn (2004). Moreover, it is available for drug expenditure data and for mortality data used as a robustness check.

1997; Jaffe 1986; Levin et al. 1985). In this paper, I follow a different approach as proposed by Andersen (1998, 1999) in which the rate of growth of a knowledge stock serves as a proxy for the technological opportunities in a particular technology area. Using this operationalization, high growth rates are associated with high technological opportunities.

The knowledge stock consists of the scientific publications (Pub_{it}) related to medical indication i and published in year t that are assigned to categories related to pharmaceutical research. Publication data was obtained in the following way. A list of 168 medical indications was drawn from the BioPharmInsight database.⁸ This list encompasses all medical indications that have been available in this database and that are not injuries or umbrella terms for several indications included in the list. Each indication describes a condition or disease that allows of the development of a drug. These medical indications were used to conduct a keyword search in the Web of Science databases (WoS) in order to gather publication data. Publications that contain at least one medical indication in the title have been taken into account. Moreover, I restrict the sample to publications that are related to pharmaceutical research, i.e., that have been assigned to one of the categories “Biochemistry & Molecular Biology”, “Biotechnology & Applied Microbiology”, “Chemistry, Applied”, “Chemistry, Medicinal”, “Medicine, Research & Experimental”, “Pharmacology & Pharmacy”, or “Toxicology”.⁹ I exclude all publications that are not labeled as journal articles. For the period from 1974 to 2008, I obtained 137,535 articles.¹⁰ Consequently, the sample includes journal articles that refer to basic and clinical pharmaceutical research. While it is evident that basic research articles reflect the development of technological opportunities through basic research’s impact on the breadth and depth of the state of knowledge, it might be less obvious to what extent clinical research contributes to the development of technological opportunities. However, Azoulay (2004) points out that knowledge production is a key aspect of clinical research activities. Particularly early phase clinical research generates new knowledge since it is dedicated to hypothesis generation and testing implying a contribution to the development of technological opportunities.¹¹

⁸Table 6 provides a list of medical indications and the corresponding therapeutic areas. The use of National Drug Codes as in Acemoglu and Linn (2004) has not been possible since the FDA does not make Major and Minor Drug Class codes available any more (FDA, 2011). The BioPharmInsight database can be found at <http://www.infinata.com/biopharma-solution/by-product/biopharm-insight.html>.

⁹A description of these categories can be found at <http://ip-science.thomsonreuters.com/cgi-bin/jmlst/jlsubcatg.cgi?PC=D>.

¹⁰Within the pharmaceutical industry articles in scientific journals are strongly connected to public and private R&D efforts, since not only universities and public research institutes but also pharmaceutical and biotechnology companies engage in intense publication activities (Godin 1996; Narin and Rozek 1988).

¹¹There may be concerns that the publication data are too narrowly defined due to the concentration on publications related to pharmaceutical research. However, case studies comparing the technological opportunities measures used in this paper and technological opportunities measures calculated based on publications in all WoS categories reveal that both measures show a rather similar development. Moreover, the case studies reveal that the technological opportunities measure responds to scientific advances related to a particular medical indication. For example, technological opportunities in the medical indication “Hepatitis” increase considerably after the isolation of the Hepatitis C virus and advances in the sequencing of the Hepatitis A virus’ genome by the end of the 1980s.

Knowledge stocks are created using the perpetual inventory method. The basic idea can be expressed as in Bitzer and Stephan (2007):

$$K_{it} = \lambda_0 Pub_{i,t} + \lambda_1 Pub_{i,t-1} + \dots + \lambda_T Pub_{i,t-T} \quad \text{with } 0 < \lambda \leq 1. \quad (4)$$

Here, λ denotes the share of knowledge of the corresponding vintage that is still in use, and T is the age of the oldest vintage of knowledge still used in t . Assuming geometric depreciation of knowledge, Eq. 4 can be transformed into

$$K_{it} = Pub_{it} + (1 - \delta) K_{i,t-1} \quad \text{with } \delta = \frac{\lambda_{t-1} - \lambda_t}{\lambda_{t-1}}, \quad (5)$$

where δ is the depreciation rate of the knowledge stock. Following previous work by Cockburn and Henderson (2001), I set δ to 20%.

According to Andersen (1998, 1999), technological opportunities can be proxied by the rate of growth of the knowledge stock. More formally, technological opportunities can be expressed as

$$TO_{it} = \frac{K_{i,t} - K_{i,t-1}}{K_{i,t-1}} * 100. \quad (6)$$

4.5 Regulatory stringency

Since the vast majority of new drugs have to be approved by the regulatory authorities, such as the FDA in the United States, the strictness of these authorities influences the number of new drugs that are approved for marketing. Regulatory stringency (*ApprovalLength_j*) is taken into account by calculating the average number of days needed for FDA approval of a new drug in the therapeutic area j , i.e., a broader disease category to which a new drug belongs. More precisely, this is the average time between the submission of a new drug approval to the FDA and its final approval (cf., Grabowski et al. 1978). The data was obtained from the Drugs@FDA database which contains information concerning the original drug application.

4.6 Pre-sample variables

Following the empirical strategy described in Section 3, pre-sample means were constructed in order to account for unobserved heterogeneity among medical indications. The pre-sample mean of new pharmaceuticals (\overline{N}_{ip}) is the average number of NMEs or NDAs approved by the FDA from 1940 to 1983 per year:

$$\overline{N}_{ip} = \frac{\sum_{t=1940}^{T=1983} N_{it}}{T - t}. \quad (7)$$

Another source of unobserved heterogeneity may be differences in the prevalence of medial indications. One may account for this aspect by introducing pre-sample market size into the analysis. However, this strategy may bias the results, since pre-sample market size is rather highly correlated with the market size variable.

Therefore, I use the mortality rate per medical indication in 1983 to account for differences in the pre-sample prevalence of medical indications:

$$M_{ip} = \frac{\sum_a Death_{it}}{\sum_a pop_{at}} * 100,000 \text{ with } t = 1983. \quad (8)$$

Pre-sample technological opportunities are constructed as the average annual growth rate of the knowledge stock from 1979 to 1983:

$$\overline{TO}_{ip} = \left(\left(\frac{K_{it=1983}}{K_{it=1979}} \right)^{1/5} - 1 \right) * 100 \quad (9)$$

5 Empirical results

5.1 Determinants of pharmaceutical innovation

Table 1 presents the results of the Poisson QML estimation without taking lagged variables into account. In models (1) and (4), I analyze the relationship between market size (MS_{it}), technological opportunities (TO_{it}), and the number of pharmaceutical innovations, i.e., NME_{it} or NDA_{it} . I find a positive and significant response of the number of new drugs to market size. This result is robust against the introduction of technological opportunities in models (3) and (6). Since market size enters in logarithmic form, we can interpret the coefficient as an elasticity. Consequently, the results presented in Table 1 suggest that a 1% increase in the market size leads to an increase in the number of NME and in the number of NDA by approximately 0.3%. This is considerably smaller than found by some other studies. Some of these studies use, however, more aggregated levels of analysis which may be a reason for these differences. While this paper uses medical indications, Acemoglu and Linn (2004) use broader categories of medications, e.g., antibiotics, that can be used as treatments for several medical indications and find a response rate of 6% to increases in the market size. Toole (2012) finds a range of the elasticity for market size between 6.8 and 12.7 using six therapeutic markets each of which may include several medical indications. In contrast to these studies, Cerda (2007) uses 15 different drug categories to examine the effect of market size on the change in the stock of drugs. The magnitude of logarithmic market size ranges from 0.11 to 0.46, which is rather similar to the magnitude of coefficients for market size in this study.

As to technological opportunities, I find a positive and significant relation to NME and NDA. The coefficient remains significant also in models (3) and (6), in which the market size is included as well. These results suggest that the number of NME and NDA responds not only to demand- but also supply-side variables. However, the coefficient is rather small, which may be a consequence of the operationalization as growth rates of knowledge stocks. Regulatory stringency ($ApprovalLength_j$) is not significantly related to the number of new pharmaceuticals. The estimates for pre-sample mortality M_{ip} do not indicate a significant relation to the number of NME

Table 1 Determinants of pharmaceutical innovation

| Dependent Variable | (1) <i>NME_{it}</i> | (2) <i>NME_{it}</i> | (3) <i>NME_{it}</i> | (4) <i>NDA_{it}</i> | (5) <i>NDA_{it}</i> | (6) <i>NDA_{it}</i> |
|-----------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| <i>MKS_{it}</i> | 0.3283*** (0.0557) | | 0.3322*** (0.0551) | 0.3138*** (0.0480) | | 0.3163*** (0.0472) |
| <i>TO_{it}</i> | | 0.0002** (0.0001) | 0.0003** (0.0001) | | 0.0002*** (0.0001) | 0.0002*** (0.0001) |
| <i>ApprovalLength_j</i> | -0.0009* (0.0005) | 0.0006 (0.0007) | -0.0008 (0.0006) | -0.0008* (0.0005) | 0.0005 (0.0006) | -0.0008* (0.0005) |
| <i>M_{ip}</i> | 0.0028 (0.0024) | 0.0066*** (0.0026) | 0.0031 (0.0024) | -0.0031 (0.0031) | 0.0006 (0.0032) | -0.0030 (0.0031) |
| \overline{TO}_{ip} | -0.0022 (0.0056) | -0.0010 (0.0049) | -0.0025 (0.0057) | -0.0019 (0.0061) | 0.0008 (0.0051) | -0.0019 (0.0061) |
| \overline{NME}_{ip} | 0.5408*** (0.1417) | 0.4094** (0.1629) | 0.5456*** (0.1418) | | | |
| \overline{NDA}_{ip} | | | | 0.6349*** (0.1068) | 0.4961*** (0.1565) | 0.6397*** (0.1037) |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -8.6394*** (1.1232) | -3.1062*** (0.4714) | -8.7496*** (1.1059) | -7.2164*** (0.8619) | -1.8144*** (0.3644) | -7.2726*** (0.8491) |
| N | 4200 | 3991 | 3991 | 4200 | 3991 | 3991 |
| Indications | 168 | 168 | 168 | 168 | 168 | 168 |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses

All market size measures enter in logarithmic form into the analysis

and NDA. Pre-sample technological opportunities \overline{TO}_{ip} show no significant effect on NME or NDA.

Pre-sample drug innovations proxied by the average number of NME (\overline{NME}_{ip}) and NDA (\overline{NDA}_{ip}) prior to the analyzed period are positively and significantly related to the number of new pharmaceuticals. This finding suggest that those medical indications for which a high number of new drugs were introduced in the past still appear attractive for innovation. Past innovations may indicate high profit margins, which make it attractive to introduce novelties in these markets to ensure current and future profits. Another possible explanation may be that a large number of past innovations shows that there has been a large amount of research in the corresponding medical indications, which may build the basis for later discoveries and further exploitation of technological opportunities in subsequent years. The positive response to pre-sample innovations contradicts the empirical evidence presented by Toole (2012), who finds a negative response at a much more aggregated level.

5.2 Time structure

So far, the analysis does not take into account the specific time structure of the innovation process in the pharmaceutical industry. Subsequent to the discovery of a new promising compound, a considerable period of time is needed in order to pass all stages of pre-clinical and clinical testing and obtain regulatory approval. On average, it takes slightly more than 7.5 years from the first clinical testing to marketing approval. The approval phase itself accounts for around 1.5 years out of this timespan. Before a new compound can enter clinical testing, researchers spend almost five years on pre-clinical investigations. Hence, on average, it takes 12.5 years from the discovery of a promising new compound to its approval (DiMasi and Grabowski 2007).

In Table 2, I use time lags in order to account for delays caused by the specificities of the industry's innovation process. More precisely, I introduce a time lag of seven years to the market size (models (1), (2), (5), (6), (7), and (10)) and technological opportunities variables (models (3), (4), (5), (8), (9), and (10)). This corresponds to the time needed on average from clinical testing to approval.

Models (1), (5), (6) and (10) show a significant and positive relationship between the lagged market size and the number of NME and NDA. In models (2) and (7), in which contemporary and lagged market size are introduced simultaneously, the coefficients of the lagged market size turn out to be negative. However, contemporary and lagged market size are not significantly related to the number of NME and NDA in this case. These findings may indicate that market size when the project started, as well as the market size when the new drug is introduced to the market, play a role in pharmaceutical R&D decisions. The positive results for lagged market size may indicate that firms base their R&D decisions at least to some extent on forecasts of the market size for the years in which they expect the new drug to enter the market. However, the estimation results do not allow for any clear-cut conclusion as to which market size measure dominates the decision.

Regarding technological opportunities, the results suggest differences between NME and NDA. The coefficients of $T O_{it-7}$ are positive, but not significantly associated to the number of NME. With an exception in model (10) the results suggest that, in case of NDA as dependent variable, contemporaneous and lagged technological opportunities are significantly and positively related to the number of new pharmaceuticals. This finding may indicate that the development of NDA builds upon a slightly different knowledge base than the more innovative NME. Past technological opportunities seem to be less important for the development of the more innovative NME. In general, the results for the control variables are comparable to those presented in Table 1.

In order to account also for the pre-clinical phase, I use 12 year time lags for market size and technological opportunities.¹² The results are rather similar to those presented in Table 2. However, lagged technological opportunities have a non-significant coefficient. Hence, technological opportunities at the beginning of the

¹²The results can be found in [Appendix](#).

Table 2 Determinants of pharmaceutical innovation with time lags

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|-----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} |
| MKS_{it} | | 1.3370 (1.0049) | 0.3444*** (0.0536) | 0.3457*** (0.0534) | | | 0.7309 (0.6116) | 0.3514*** (0.0495) | 0.3521*** (0.0494) | |
| MKS_{it-7} | | 0.3375*** (0.0533) | -1.0035 (1.0056) | | 0.3442*** (0.0540) | 0.3474*** (0.0493) | -0.3847 (0.6108) | | | 0.3510*** (0.0496) |
| TO_{it} | | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0002*** (0.0001) | 0.0002*** (0.0001) | 0.0002*** (0.0001) | | 0.0002*** (0.0001) | |
| TO_{it-7} | | | 0.0002 (0.0002) | 0.0002 (0.0002) | 0.0002 (0.0002) | | | 0.0002* (0.0001) | 0.0002* (0.0001) | 0.0002 (0.0001) |
| $ApprovalLength_j$ | | -0.0011* (0.0006) | -0.0012* (0.0006) | -0.0012* (0.0006) | -0.0012* (0.0006) | -0.0009* (0.0005) | -0.0009* (0.0005) | -0.0009* (0.0005) | -0.0009* (0.0005) | -0.0009* (0.0005) |
| M_{ip} | | 0.0032 (0.0027) | 0.0029 (0.0027) | 0.0032 (0.0030) | 0.0032 (0.0030) | -0.0036 (0.0032) | -0.0039 (0.0032) | -0.0046 (0.0038) | -0.0046 (0.0038) | -0.0044 (0.0037) |
| \overline{TO}_{ip} | | -0.0022 (0.0064) | -0.0027 (0.0065) | -0.0023 (0.0066) | -0.0021 (0.0065) | -0.0015 (0.0059) | -0.0018 (0.0059) | -0.0015 (0.0058) | -0.0014 (0.0058) | -0.0013 (0.0058) |
| \overline{NME}_{ip} | | 0.5340*** (0.1395) | 0.5428*** (0.1392) | 0.5315*** (0.1401) | 0.5354*** (0.1404) | 0.5291*** (0.1404) | | | | |
| \overline{NDA}_{ip} | | | | | | | | | | |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -8.6722*** (1.0429) | -8.7788*** (1.0452) | -8.7829*** (1.0664) | -8.8204*** (1.0627) | -8.7371*** (1.0672) | -7.7737*** (0.8251) | -7.8400*** (0.8288) | -7.8847*** (0.8325) | -7.9121*** (0.8288) | -7.8395*** (0.8296) |

Table 2 (continued)

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|--------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} |
| N | 2964 | 2964 | 2817 | 2817 | 2817 | 2964 | 2964 | 2817 | 2817 | 2817 |
| Indications | 168 | 168 | 167 | 167 | 167 | 168 | 168 | 167 | 167 | 167 |
| Standard errors in parentheses | | | | | | | | | | |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses

All market size measures enter in logarithmic form into the analysis

pre-clinical research phase are not significantly related to the number of NME and NDA. A possible explanation for this finding may be found in the high uncertainty accompanying pharmaceutical R&D. High technological opportunities at the beginning of the pre-clinical research phase may lead to extensive exploration, but most of the possible compounds do not obtain market approval. In contrast, a high level of technological opportunities at the beginning of the clinical research phase may enable pharmaceutical companies to improve promising compounds that may have a higher probability of finally obtaining market approval.

5.3 Clinical trials

In order to examine further the robustness of the relationship between market size, technological opportunities and new pharmaceuticals in the clinical phases of the R&D process, I use the number of clinical trials in phase II and in phase III as dependent variables. These dependent variables serve as proxies for the number of new compounds under development. The focus on clinical testing requires different lag structures as compared to the lags used in the regressions using NMEs and NDAs as dependent variables. Following DiMasi and Grabowski (2007), drug candidates need around one year on average to enter phase II clinical trials. Therefore, I use a time lag of one year for market size and technological opportunities in the models in Table 3 in which the number of phase II clinical trials is the dependent variable. With respect to phase III clinical trials, I use a time lag of three years for market size and technological opportunities. This time lag corresponds to the approximately three years of clinical testing new compounds need on average to enter phase III clinical trials (see, DiMasi and Grabowski 2007).

As in the previous regressions, the coefficients for market size and lagged market size are positive and significant. Exceptions can be found in models (3) and (9) in Table 3, in which market size is introduced together with lagged market size. The results for contemporaneous technological opportunities do not indicate a robust significant relation to the number of clinical trials. However, with an exception in model (6), lagged technological opportunities are positively and significantly related to the number of phase II and phase III clinical trials. These results suggest that the number of clinical trials responds to both market size and (lagged) technological opportunities. Hence, supply side and demand side variables might increase the number of new compounds that are tested and might be approved as new drugs in the future.

The results of the control variables differ partly from those obtained in the previous analyses. Regulatory stringency ($ApprovalLength_j$) is positively related to the number of phase III clinical trials. A possible explanation for this finding might be that the regulatory authorities demand further phase III clinical tests during the examination of new compounds. These additional clinical studies should provide further evidence for the safety and efficacy of a new compound. In contrast, the number of phase II clinical trials may be more directly influenced by demand and supply side factors as indicated by the significant coefficients for the prevalence of medical indications and pre-sample technological opportunities. The pre-sample mean of new drug approvals is positively related to phase II and phase III clinical trials. As in the case for NMEs and NDAs, these findings may indicate that medical indications

Table 3 Determinants of clinical trials

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|-----------------------|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | $Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}, Ph2Trials_{it}$ | | | | | | | | | | | |
| MKS_{it} | 0.2540*** (0.0882) | 0.9848* (0.5892) | 0.2564*** (0.0883) | 0.2558*** (0.0883) | 0.4657*** (0.0949) | 0.6055 (1.4070) | 0.4665*** (0.0962) | 0.6055 (1.4070) | 0.4699*** (0.0953) | 0.4700*** (0.0953) | | |
| MKS_{it-1} | | 0.2506*** (0.0884) | -0.7307 (0.5718) | 0.4674*** (0.0950) | | | | | | | | |
| MKS_{it-3} | | | | | | | 0.4665*** (0.0962) | -0.1405 (1.4099) | | | | 0.4705*** (0.0966) |
| TO_{it} | -0.0002 (0.0002) | -0.0002 (0.0001) | -0.0002 (0.0002) | -0.0002* (0.0001) | -0.0001 (0.0003) | -0.0001 (0.0003) | -0.0001 (0.0003) | -0.0001 (0.0003) | 0.0000 (0.0001) | | | |
| TO_{it-1} | | | | 0.0001*** (0.0000) | 0.0001*** (0.0000) | -0.0001 (0.0003) | | | | | | |
| TO_{it-3} | | | | | | | | | | 0.0003*** (0.0000) | 0.0003*** (0.0000) | 0.0003*** (0.0000) |
| $ApprovalLength_{it}$ | -0.0008 (0.0008) | -0.0008 (0.0008) | -0.0007 (0.0008) | -0.0008 (0.0008) | -0.0008 (0.0008) | 0.0023** (0.0010) | 0.0023** (0.0010) | 0.0024** (0.0010) | 0.0024** (0.0010) | 0.0024** (0.0010) | 0.0024** (0.0010) | 0.0024** (0.0010) |
| M_{it} | 0.0056* (0.0031) | 0.0057* (0.0031) | 0.0055* (0.0031) | 0.0056* (0.0031) | 0.0056* (0.0031) | -0.0036 (0.0045) | -0.0037 (0.0045) | -0.0038 (0.0045) | -0.0036 (0.0045) | -0.0036 (0.0045) | -0.0036 (0.0045) | -0.0032 (0.0044) |
| \overline{TO}_{it} | 0.0131* (0.0071) | 0.0131* (0.0071) | 0.0129* (0.0071) | 0.0132* (0.0071) | 0.0132* (0.0071) | 0.0016 (0.0068) | 0.0014 (0.0068) | 0.0016 (0.0068) | 0.0018 (0.0068) | 0.0018 (0.0068) | 0.0018 (0.0068) | 0.0021 (0.0068) |
| \overline{NDA}_{it} | 0.5474*** (0.0000) | 0.5473*** (0.0000) | 0.5512*** (0.0000) | 0.5515*** (0.0000) | 0.5502*** (0.0000) | 0.7157*** (0.0000) | 0.7173*** (0.0000) | 0.7206*** (0.0000) | 0.7171*** (0.0000) | 0.7258*** (0.0000) | 0.7261*** (0.0000) | 0.7295*** (0.0000) |

Table 3 (continued)

| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|--------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Dependent Variable | $Ph2Trials_{it}$ | $Ph2Trials_{it}$ | $Ph2Trials_{it}$ | $Ph2Trials_{it}$ | $Ph2Trials_{it}$ | $Ph2Trials_{it}$ | $Ph2Trials_{it}$ | $Ph3Trials_{it}$ | $Ph3Trials_{it}$ | $Ph3Trials_{it}$ | $Ph3Trials_{it}$ | $Ph3Trials_{it}$ |
| Year | (0.1425) | (0.1432) | (0.1427) | (0.1426) | (0.1426) | (0.1336) | (0.1338) | (0.1341) | (0.1333) | (0.1342) | (0.1342) | (0.1345) |
| Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -3.4790* | -3.4292* | -3.5439* | -3.5398* | -3.5223* | -10.0006*** | -9.9965*** | -9.9851*** | -9.9987*** | -10.1187*** | -10.1228*** | -10.1078*** |
| | (1.8446) | (1.8492) | (1.8486) | (1.8496) | (1.8492) | (1.7526) | (1.7496) | (1.7658) | (1.7499) | (1.7594) | (1.7602) | (1.7761) |
| N | 1508 | 1508 | 1508 | 1507 | 1507 | 1507 | 1508 | 1508 | 1508 | 1502 | 1502 | 1502 |
| Indications | 168 | 168 | 168 | 168 | 168 | 168 | 168 | 168 | 168 | 168 | 168 | 168 |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses

All market size measures enter in logarithmic form into the analysis

for which successful innovation took place in the past remain to be attractive and profitable markets.

Although new drug approvals are related to clinical trials, since clinical testing is a precondition for approval by the regulatory authorities, the number of clinical trials may be influenced by different mechanisms than the number of approved drugs. In order to account for this aspect I use pre-sample means for phase II as well as for phase III clinical trials based on the years 2000 to 2002 to account for unobserved heterogeneity in the clinical research activities among different medical indications. The corresponding results reveal a rather robust positive relationship between (lagged) market size and the number of clinical trials. Furthermore, lagged technological opportunities show a rather robust positive relationship to the number of phase II and phase III trials, whereas no robust positive relationship for contemporary technological opportunities has been found.

5.4 Robustness checks

Table 4 presents several robustness checks. As pointed out by Acemoglu and Linn (2004), changes in the insurance coverage of drug expenditures can be another source of market size variation. In order to proxy drug expenditure coverage, I multiply the market size measure by the fraction of privately insured persons in age group a at time t , (ins_{at}). The data were obtained from the Medical Expenditure Panel Survey (MEPS), a large scale survey of U.S. families and individuals, their medical providers, and employers. Equation 3 can be modified to $\widetilde{MKS}_{it} = \sum_a \omega_{ia} pop_{at} inc_{at} ins_{at}$. Models (1) and (7) present the regression analysis using \widetilde{MKS}_{it} . The results are rather similar to the results presented above. As in Table 1, I find a positive and significant response for NME and NDA to both market size and technological opportunities. The results stay qualitatively the same if the insurance coverage also encompasses other types of health insurance.

In models (2) and (8) in Table 4, market size is constructed using weights representing the average fraction of individuals dying due to a specific indication divided by the total number of deaths in an age group for the years 1996 to 1998 (cf., Cerda 2007). Again, the weights are not time varying, in order to rule out changes in the use of drugs for a certain medical indication due to price and quality changes as well as the introduction of new pharmaceuticals. The data on multiple causes of death in the United States were obtained from the National Vital Statistics System of the National Center of Health Statistics. The results suggest a positive but not significant relation between market size and the number of new pharmaceuticals. Technological opportunities are again positively associated with the number of NME and NDA.

In (3) and (9), technological opportunities are constructed by only taking biotechnology publications into account. These are journal articles assigned to the categories “Biochemistry & Molecular Biology” and “Biotechnology & Applied Microbiology” in the WoS databases. Whereas the market size variable is still significantly positive related to the number of NME and NDA in this setting, the coefficient for technological opportunities is not significant. In models (4) and (10), I employ only basic research articles to construct technological opportunities using the CHI

Table 4 Robustness checks

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|-----------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | <i>NME_{it}</i> | <i>NME_{it}</i> | <i>NME_{it}</i> | <i>NME_{it}</i> | <i>NME_{it}</i> | <i>NME_{it}</i> | <i>NDA_{it}</i> | <i>NDA_{it}</i> | <i>NDA_{it}</i> | <i>NDA_{it}</i> | <i>NDA_{it}</i> | <i>NDA_{it}</i> |
| <i>MKS_{it}</i> | 0.3255*** (0.0486) | 0.0548 (0.0524) | 0.3367*** (0.0578) | 0.3293*** (0.0542) | 0.3251*** (0.0521) | 1.2339 (1.1624) | 0.3452*** (0.0493) | 0.0035 (0.0370) | 0.3299*** (0.0451) | 0.3119*** (0.0463) | 0.3147*** (0.0492) | 1.4736 (1.0550) |
| <i>TO_{it}</i> | 0.0003*** (0.0001) | 0.0002** (0.0001) | -0.0000 (0.0003) | 0.0008** (0.0003) | 0.0013*** (0.0005) | 0.0008*** (0.0002) | 0.0003*** (0.0001) | 0.0002*** (0.0001) | -0.0000 (0.0002) | 0.0005 (0.0003) | 0.0007*** (0.0003) | 0.0005*** (0.0001) |
| <i>ApprovalLength_j</i> | -0.0013* (0.0007) | 0.0005 (0.0007) | -0.0008 (0.0006) | -0.0010* (0.0006) | -0.0009 (0.0006) | | -0.0008 (0.0005) | 0.0005 (0.0006) | -0.0005 (0.0005) | -0.0009* (0.0005) | -0.0007 (0.0005) | |
| <i>M_{ip}</i> | 0.0007 (0.0031) | 0.0040 (0.0032) | 0.0026 (0.0025) | 0.0028 (0.0024) | 0.0026 (0.0027) | | -0.0046 (0.0033) | 0.0004 (0.0034) | -0.0035 (0.0033) | -0.0034 (0.0032) | -0.0031 (0.0032) | |
| <i>TO_{ip}</i> | -0.0036 (0.0081) | -0.0023 (0.0048) | 0.0028 (0.0073) | -0.0076 (0.0054) | -0.0088 (0.0101) | | -0.0033 (0.0072) | 0.0008 (0.0051) | 0.0048 (0.0060) | -0.0025 (0.0063) | -0.0336*** (0.0128) | |
| <i>NME_{ip}</i> | 0.5088*** (0.1333) | 0.4314*** (0.1621) | 0.5182*** (0.1475) | 0.5308*** (0.1413) | 0.5368*** (0.1404) | | | | | | | |
| <i>NDA_{ip}</i> | | | | | | | | | | | | |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -8.1387*** (0.8728) | -4.2349*** (1.3266) | -8.8426*** (1.1289) | -8.5296*** (1.0791) | -8.5171*** (1.0305) | | 0.6251*** (0.1269) | 0.4976** (0.1540) | 0.6489*** (0.0987) | 0.6320*** (0.1014) | 0.6380*** (0.1207) | -7.1019*** (0.8364) |
| | | | | | | | | | | | | -7.0928*** (0.8510) |

Table 4 (continued)

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|--------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} |
| N | 2169 | 3987 | 3112 | 3784 | 3088 | 3174 | 2169 | 3987 | 3112 | 3784 | 3088 | 3426 |
| Indications | 168 | 168 | 149 | 165 | 159 | 132 | 168 | 168 | 149 | 165 | 159 | 143 |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses

All market size measures enter in logarithmic form into the analysis

In models (1) and (7) market size is multiplied with the share of privately insured individuals in an age group at t ,

(2) and (8) market size is constructed using weights representing the average share of individuals dying because of i ,

(3) and (9) take only biotechnology publications into account when constructing technological opportunities,

(4) and (10) only basic research publications into account when constructing technological opportunities,

(5) and (11) use an 10% depreciation rate of the knowledge stock used for TO_{it} ,

(6) and (12) use Poisson QML estimation with fixed effects

classification of journals (Hamilton 2003). The CHI classification categorizes each article according to the type of research prevalent in the journal in which it was published. Taking only “basic biomedical research” into account, the response of the number of new pharmaceuticals to market size remains significantly positive whereas the coefficient for technological opportunities is not significant for NDA in model (10). The non-significance of basic research and biotechnology-related technological opportunities may be caused by the construction of the publication database. Since a keyword search for medical indications was used, the data may capture mainly applied research and only a few basic and biotechnology research papers building the basis for new pharmaceuticals for several medical indications. In particular, the data may not include new research methods that encompass a considerable amount of innovation in biotechnology and may be more indirect in how they affect the number of new pharmaceuticals. Moreover, the relevant knowledge stocks for basic research may differ from the ones for more applied research.

In (5) and (11) I use a 10% depreciation rate of the knowledge stock when constructing the proxy for technological opportunities. The coefficients for technological opportunities are positive and significant for NME and NDA. Hence, the technological opportunities measure seems to be robust against changes in the depreciation rate. Market size is still positively and significantly related to the number of new pharmaceuticals. In regression models (6) and (12) I use a Poisson QML estimation with fixed effects for medical indications. In these models, the market size measure is not significant. This result is rather surprising but may be the consequence of the fixed effects model. Since I run the analysis on a rather disaggregated level, there are a considerable number of medical indications for which no NME or NDA was introduced over the entire period of analysis. These indications are not taken into account in the fixed effects estimation, and this may have consequences for the results. Nevertheless, the proxy for technological opportunities is significantly positively related to the number of NME and NDA. Moreover, the results presented in Table 1 are generally robust to the use of ten year age groups in the construction of the market size proxy and linear depreciation of the knowledge stock over a five year period.

I account for the possibility that the relation between market size and the number of new pharmaceuticals depends on the level of technological opportunities related to specific medical indications by including interaction terms between market size and technological opportunities into the analysis. The corresponding results do neither suggest a statistically significant interaction between contemporary market size and contemporary technological opportunities nor a significant interaction between lagged market size and lagged technological opportunities using lags of seven years. Following Blundell et al. (1995) a dummy variable accounting for zero pre-sample new pharmaceuticals for indication i has been included in the analysis. The corresponding results are qualitatively similar to those presented above. However, the technological opportunities measure loses its significance in the models using NDA_{it} as dependent variable when time lags of seven years and the zero pre-sample new pharmaceuticals dummy are used.

There might be concerns that the subsample of publications containing a medical indication in their title is too narrowly defined to account for the scientific developments that influence technological opportunities corresponding to specific medical

indications. Therefore, I calculate technological opportunities based on an extended sample of scientific publications in order to account for a broader set of indication-specific research activities. More specifically, I calculate technological opportunities using 828,280 publications that contain at least one medical indication in their title, abstract, or keywords and that have been assigned to one of the categories listed in Section 4.4. The corresponding results are qualitatively similar to those presented above.

With respect to the computation of the technological opportunities measure, further concerns may refer to the development of the number of relevant scientific journals and their coverage by the WoS. The number of relevant scientific journals has been growing over the past decades and the coverage of articles and journals by the WoS may have changed over time, too. In addition, the information accessible to researchers may have changed over time. For instance, the WoS claims to provide searchable "full-length, English-language author abstracts" for approximately 70% of the articles from 1991 onwards. Consequently, there might be concerns that the computation of the technological opportunities measure is influenced by changes in the WoS database. In order to account for these changes in the database, in addition to the use of year dummies in the regression analyses, I conduct a robustness check taking only scientific articles published after 1991 into account to calculate the technological opportunities measure. The corresponding results are qualitatively similar to those presented above.

Since the results may depend on the level of analysis, namely medical indications, a robustness check using the more aggregated level of therapeutic areas that encompass several medical indications is conducted, too. The results are qualitatively similar to the results on the medical indication level.

A further concern might be that larger markets attract more public or private research, generating higher technological opportunities. One may check for these feedback effects by taking only rare diseases into account, since there are only a few market incentives to develop medications for them. However, the number of rare diseases in the sample is quite small and the corresponding Poisson models do not converge. Moreover, one may think about differences in the relationship between market size and pharmaceutical innovation between the U.S. and other important markets such as Europe and Japan. However, detailed data on drug spendings, population numbers per age group, and the number of new drugs corresponding to the indications in order to address this issue have not been available to the author.

5.5 Reverse causality

As pointed out by Kleinknecht and Verspagen (1990), there may be a mutual dependency between market size and innovation instead of a unidirectional causality. Particularly in the pharmaceuticals industry, new products may introduce the potential for reverse causality. Lichtenberg (2004, 2005b) shows that new drugs contributed to the longevity increase observed in many countries over the past few decades. Hence, through their impact on consumers' life expectancy, new drugs lead to an endogenous growth of pharmaceutical markets (Cerda 2007). However, Acemoglu and Linn (2004) argue that the drug-induced changes in the population sizes are quite

small relative to overall demographic changes. Nevertheless, they propose a strategy to address this issue. Following their approach, I instrument current market size by the corresponding market size five years earlier. More precisely, I use the size and income of those 70–74 years old in 1984 to instrument for population size and income of those 75–79 years old in 1989. The population size and the income of those 70–74 years old is highly correlated with the size and the income of those 75–79 years old five years later, but it is unaffected by new pharmaceuticals that were approved in the meantime.

In the empirical analysis of this instrumental-variable I use a Generalized Methods of Moments (GMM) estimator for the pooled Poisson regression. Standard errors are bootstrapped with 1000 replications and clustered on the indication level. The results are presented in Table 5. These results are quite similar to the baseline results in Table 1. The number of new NME and NDA responds positively to market size. However, the coefficients for technological opportunities are not statistically significant.

Table 5 Determinants of pharmaceutical innovation GMM estimation

| Dependent Variable | (1) <i>NME_{it}</i> | (2) <i>NME_{it}</i> | (3) <i>NDA_{it}</i> | (4) <i>NDA_{it}</i> |
|-----------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| <i>MKS_{it}</i> | 0.2451*** (0.0609) | 0.2502*** (0.0685) | 0.1915*** (0.0589) | 0.2063*** (0.0599) |
| <i>TO_{it}</i> | | 0.0001 (0.0052) | | 0.0001 (0.0014) |
| <i>ApprovalLength_j</i> | -0.0004 (0.0007) | -0.0004 (0.0007) | -0.0003 (0.0006) | -0.0002 (0.0006) |
| <i>M_{ip}</i> | -0.0002 (0.0064) | 0.0011 (0.0088) | -0.0031 (0.0036) | -0.0023 (0.0037) |
| <i>TO_{ip}</i> | -0.0029 (0.0076) | -0.0029 (0.0075) | -0.0022 (0.0069) | -0.0027 (0.0070) |
| <i>NME_{ip}</i> | 1.0656*** (0.2585) | 1.0557*** (0.2656) | | |
| <i>NDA_{ip}</i> | | | 1.3572*** (0.1951) | 1.3190*** (0.1904) |
| Year Dummies | Yes | Yes | Yes | Yes |
| Constant | -7.4008*** (1.0537) | -7.4919*** (1.1546) | -6.0008*** (1.0407) | -6.2739*** (1.0221) |
| N | 4200 | 3991 | 4200 | 3991 |
| Indications | 168 | 168 | 168 | 168 |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses

All market size measures enter in logarithmic form into the analysis

Market size is instrumented with market size of the same age cohort five years earlier

Consequently, I find no evidence that the problem of reverse causality arises in the analysis.

6 Conclusion

Recent empirical contributions have stressed the importance of the (potential) market size for the development of new pharmaceuticals. At the same time, many scholars have emphasized the changes in the industry's R&D process, from "random screening" to "guided drug development", and have pointed to the importance of advances in molecular biology and related fields for the industry's technological opportunities and innovative capabilities. Against this background, I have analyzed the relationships between (potential) market size, technological opportunities, and the number of new pharmaceuticals in the United States. This empirical analysis is based on a unique dataset combining data from different sources. In the analysis, I use an "entry stock" Poisson quasi-maximum likelihood estimator (QMLE), as suggested by Blundell et al. (1995).

The empirical estimates reveal a significantly positive response of the number of new pharmaceuticals, i.e., new molecular entities or new drug approvals, to market size and technological opportunities. Similar results are obtained when the dependent variable is the number of phase II or phase III clinical trials. In the robustness checks, the market size measures lose significance when mortality data is used in its construction and when the PQML fixed effects model is used. As a further robustness check, I constructed the technological opportunities proxy by taking only biotechnology or basic research publications into account. The results from this indicated no significantly robust positive relation between these two variants of the technological opportunities measure and the number of new pharmaceuticals. When taking the specific time structure of the pharmaceutical R&D process into account by including time lags of seven years into the analysis, I found no clear-cut relationship between lagged market size, lagged technological opportunities, and the number of new pharmaceuticals. The results stay qualitatively similar if broader therapeutic areas instead of medical indications are used as a unit of analysis.

Overall, the results suggest a quite robust and significantly positive relation between the (potential) market size and the number of new pharmaceuticals, as well as between (potential) market size and the number of clinical trials. These results are in line with recent contributions by Cerda (2007) and Acemoglu and Linn (2004). I found no evidence of a reverse causality between market size and the number of new pharmaceuticals. Regarding technological opportunities, I found a significantly positive relation with the number of new drugs. Technological opportunities at the beginning of the clinical development phase are positively related to the number of phase II and phase III clinical trials. These results support earlier studies indicating the importance of technological opportunities in themselves and the scientific progress, which is an important source of new opportunities for pharmaceutical innovation (e.g., Toole 2012; Lim 2004; Griliches et al. 1991). Nevertheless, the results for technological opportunities need further analysis of the role of advances in biotechnology and basic research for the development of new pharmaceuticals.

Since my results suggest that the (potential) market size influences the rate and direction of innovative activities in pharmaceuticals, policy makers can support the development of drugs for medical indications with small markets through incentive creating programs, e.g., tax credits. Nevertheless, these programs have to take into account the fact that private R&D investment is rather concentrated in the bigger markets among the small ones (Yin 2008). The empirical analysis suggests that technological opportunities are positively related to the number of clinical trials as well as to the number of new pharmaceuticals. These results indicate that supply side policies can provide another promising way to stimulate the development of new pharmaceuticals and new compounds. Public funding and other support of scientific research directed to certain diseases can lead to the generation of new scientific knowledge, contributing to a renewal of the technological opportunities. The latter can be seen as a set of possibilities from which new drugs can be developed. As a consequence, the number of promising compounds may increase and translate into new pharmaceuticals after the different stages of clinical research. This process of generating new technological opportunities and translating these into new pharmaceuticals may, however, be time consuming. Following DiMasi and Grabowski (2007), drug development takes on average more than 12 years from the discovery of a promising compound to market approval. Taking the generation of new technological opportunities into account, the lag between public research funding and market approval may be close to the 17 and 24 years, as suggested by Toole (2012). Hence, public research support has to have a long-term orientation.

As shown in previous studies, public investments in pharmaceutical research can have important indirect effects increasing the industry's innovative activities through the stimulation of private R&D investment (Toole 2007). Recommendations concerning the support of different research types to increase the number of new pharmaceuticals need further evaluations of the role of biotechnology and basic research. However, research addressing the mechanism of action of drugs and the causes of diseases might be promising, since both are not yet fully understood. Increased knowledge in these areas may help to discover and to synthesize new compounds with improved safety and efficacy. These quality improvements may help to reduce the cost of the clinical phases of the pharmaceutical R&D process. Regarding regulatory stringency, the empirical analysis does not suggest that reductions in this stringency would be an appropriate method for spurring pharmaceutical innovation.

Acknowledgements I am grateful to the German Research Foundation (DFG) for financial support (DFG GKR 1411). I thank Uwe Cantner, Matthew J. Higgins, Ivan Savin, Simon Wiederhold, the participants of the Research Seminar at the Institute for Innovation and Entrepreneurship in Gothenburg, Sweden on February the 3rd 2012, the participants of the DRUID Society Conference 2012 in Copenhagen, Denmark, the participants of the 14th ISS Conference in Brisbane, Australia, the participants of the Annual Congress of the Verein für Socialpolitik 2012 in Göttingen, Germany and two anonymous reviewers for useful comments, expressed interest and concerns. The usual caveats apply.

Compliance with Ethical Standards

Conflict of interests The author declares that he has no conflict of interest.

Appendix

A.1 List of medical indications

Table 6 List of medical indications, source: BioPharmInsight

| Therapeutic Area | Indications |
|-------------------------|--|
| Cancer | Bladder Cancer, Bone Cancer, Brain Cancer, Breast Cancer, Cervical Cancer, Colon Cancer, Gastrointestinal Cancer, Head and Neck Cancer, Kidney Cancer, Leukemia, Liver Cancer, Lung Cancer, Lymphoma, Melanoma, Mucositis, Multiple Myeloma, Non Hodgkins Lymphoma, Ovarian Cancer, Pancreatic Cancer, Soft Tissue Sarcoma |
| Cardiovascular Diseases | Acute Coronary Syndrome, Angina, Arrhythmias, Atherosclerosis, Cardiovascular Disease, Heart Failure, Hypertension, Hypoxia, Myocardial Infarction, Myocardial Ischemia, Stroke, Thrombosis |
| Central Nervous System | Alcohol & Drug Dependence, Alzheimer's Disease, Ataxia, Autism, Bipolar Disorder, Chronic Fatigue Syndrome, CNS Disorders, Dementia, Eating Disorders, Epilepsy, Huntington's Disease, Insomnia, Mad Cow Disease, Multiple Sclerosis, Narcolepsy, Neurodegenerative Diseases, Neuropathies, Parkinson's Disease, Restless Leg Syndrome, Schizophrenia, Seizures & Spasms, Sleep Apnea, Spinal Cord Injuries |
| Dermatology | Acne, Dermatitis, Psoriasis, Rosacea |
| Eye and Ear | Cataracts, Dry Eye, Glaucoma, Macular Degeneration, Macular Edema, Otitis Media, Retinopathy, Uveitis |
| Gastrointestinal | Colitis, Constipation, Crohn's Disease, Diarrhea, Inflammatory Bowel Disease, Irritable Bowel Syndrome, Liver Disease, Obesity, Short Bowel Syndrome, Ulcers |
| Genitourinary | Benign Prostatic Hyperplasia, Endometriosis, Incontinence, Overactive Bladder, Polycystic Ovarian Disease, Prostate Related Diseases, Prostatitis, Sexual Dysfunction, Urethritis, Uterine Fibroids |
| Hematological | Anemia, Coagulation, Hemophilia, Iron Overload, Sickle Cell Anemia & Thalassemia, Thrombocytopenia, von Willebrand Factor |
| Hormonal Systems | Acromegaly, Diabetes, Diabetic Nephropathy, Gaucher's Disease, Hormonal Systems, Hyperlipidemia & Dyslipidemia, Hyperparathyroidism, Hyponatremia, Hypothyroidism, Kidney Disease, Menopause, Metabolic Conditions, Mucopolysaccharidoses, Nephritis, Pompe's Disease, Thyroid Disease |
| Immune System | Allergies, Amyloidosis, Autoimmune Diseases, Canker Sores, Graft-versus-host Disease, Inflammatory Diseases, Lupus, Periodontal Disease |
| Infectious Diseases | Bacterial Vaginosis, Candidiasis, Cytomegalovirus Infections, Diphtheria & Pertussis & Tetanus, Escherichia Coli, Encephalitis, Fungal Infections, Genital Herpes, Gram-negative Infections, Gram-positive Infections, Hepatitis, Herpes, Impetigo, (Other) Infectious Diseases, Influenza, Malaria, Measles & Mumps & Rubella, Meningitis, Mononucleosis, Mycobacterium Avium Complex, Pharyngitis, Pneumonia, Poisoning, Respiratory Infections, Rotavirus Infections, Respiratory Syncytial Virus Diseases, Sepsis, Septic Shock, Severe Acute Respiratory Syndrome, Sexually Transmitted Diseases, Tonsillitis, Tuberculosis, Urinary Tract Infections |

Table 6 (continued)

| Therapeutic Area | Indications |
|------------------|--|
| Musculoskeletal | Ankylosing Spondylitis, Arthritis, Muscular Dystrophy, Osteoarthritis, Osteoporosis, Paget’s Disease, Rheumatoid Arthritis |
| Pain | Migraine & Headache, Pain |
| Respiratory | Acute Respiratory Distress Syndrome, Asthma, Bronchitis, Chronic Obstructive Pulmonary Disease, Cystic Fibrosis, Emphysema, Respiratory, Sinusitis |

A.2 Descriptive statistics

Table 7 Descriptive statistics

| | | Mean | Std. Dev. | Min | Max | Observations | |
|----------------------|---------|----------|-----------|-----------|-----------|--------------|--------|
| NME_{it} | overall | 0.1557 | 0.4547 | 0.0000 | 5.0000 | N | 4200 |
| | between | | 0.2124 | 0.0000 | 1.8800 | n | 168 |
| | within | | 0.4024 | -1.7243 | 3.2757 | T | 25 |
| NDA_{it} | overall | 0.5557 | 1.1389 | 0.0000 | 12.0000 | N | 4200 |
| | between | | 0.7365 | 0.0000 | 4.7600 | n | 168 |
| | within | | 0.8706 | -3.2043 | 11.3557 | T | 25 |
| $Ph2Trials_{it}$ | overall | 3.2368 | 7.9443 | 0.0000 | 78.0000 | N | 1512 |
| | between | | 7.1187 | 0.0000 | 52.1111 | n | 168 |
| | within | | 3.5645 | -25.2077 | 43.3479 | T | 9 |
| $Ph3Trials_{it}$ | overall | 2.4537 | 5.8684 | 0.0000 | 78.0000 | N | 1512 |
| | between | | 4.8894 | 0.0000 | 41.2222 | n | 168 |
| | within | | 3.2649 | -33.7685 | 39.2315 | T | 9 |
| MKS_{it} | overall | 18.2070 | 1.9863 | 10.8872 | 23.2560 | N | 4200 |
| | between | | 1.9845 | 11.2216 | 22.9477 | n | 168 |
| | within | | 0.1733 | 17.6938 | 18.8297 | T | 25 |
| TO_{it} | overall | 14.7410 | 187.6144 | -20.0000 | 9074.9480 | N | 3991 |
| | between | | 45.2071 | -20.0000 | 487.9656 | n | 168 |
| | within | | 183.1949 | -493.2246 | 8601.7240 | T-bar | 23.756 |
| $ApprovalLength_j$ | overall | 546.8478 | 145.6267 | 307.8000 | 780.1364 | N | 4200 |
| | between | | 146.0446 | 307.8000 | 780.1364 | n | 168 |
| | within | | 0.0000 | 546.8478 | 546.8478 | T | 25 |
| M_{ip} | overall | 5.8228 | 19.5655 | 0.0004 | 127.0060 | N | 4200 |
| | between | | 19.6217 | 0.0004 | 127.0060 | n | 168 |
| | within | | 0.0000 | 5.8228 | 5.8228 | T | 25 |
| \overline{TO}_{ip} | overall | 8.8835 | 13.1197 | -16.3488 | 96.4633 | N | 4200 |
| | between | | 13.1574 | -16.3488 | 96.4633 | n | 168 |
| | within | | 0.0000 | 8.8835 | 8.8835 | T | 25 |

Table 7 (continued)

| | | Mean | Std. Dev. | Min | Max | Observations | |
|-----------------------|---------|--------|-----------|--------|--------|--------------|------|
| \overline{NME}_{ip} | overall | 0.6816 | 0.6761 | 0.0000 | 6.0000 | N | 4200 |
| | between | | 0.6781 | 0.0000 | 6.0000 | n | 168 |
| | within | | 0.0000 | 0.6816 | 0.6816 | T | 25 |
| \overline{NDA}_{ip} | overall | 0.8786 | 0.7648 | 0.0000 | 6.0000 | N | 4200 |
| | between | | 0.7670 | 0.0000 | 6.0000 | n | 168 |
| | within | | 0.0000 | 0.8786 | 0.8786 | T | 25 |

Table 8 Correlations

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---------------------------|-----------|-----------|-----------|---------|-----------|-----------|--------|
| (1) MKS_{it} | 1.0000 | | | | | | |
| (2) TO_{it} | -0.0363* | 1.0000 | | | | | |
| (3) $ApprovalLength_j$ | 0.2805*** | -0.0282 | 1.0000 | | | | |
| (4) M_{ip} | 0.2022*** | -0.0058 | 0.1625*** | 1.0000 | | | |
| (5) \overline{TO}_{ip} | 0.1047*** | -0.0501** | 0.0467** | 0.0323* | 1.0000 | | |
| (6) \overline{NME}_{ip} | 0.0928*** | -0.0315* | -0.0339* | 0.0279 | 0.0778*** | 1.0000 | |
| (7) \overline{NDA}_{ip} | 0.1664*** | -0.0274 | -0.0230 | -0.0174 | 0.0673*** | 0.9028*** | 1.0000 |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

A.3 Results with 12-year time lags

Table 9 Determinants of pharmaceutical innovation with 12-year time lags

| Dependent Variable | (1) NME_{it} | (2) NME_{it} | (3) NME_{it} | (4) NME_{it} | (5) NME_{it} | (6) NDA_{it} | (7) NDA_{it} | (8) NDA_{it} | (9) NDA_{it} | (10) NDA_{it} |
|-----------------------|-----------------------|-----------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| MKS_{it} | 0.9085 (1.0001) | 0.9085 (1.0001) | 0.3413*** (0.0499) | 0.3436*** (0.0496) | 0.3411*** (0.0504) | 0.3464*** (0.0500) | 0.7943 (0.5178) | 0.3584*** (0.0501) | 0.3599*** (0.0499) | 0.3580*** (0.0503) |
| MKS_{it-12} | 0.3269*** (0.0492) | -0.5845 (1.0009) | | | | | | | | |
| TO_{it} | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) | 0.0003*** (0.0001) |
| TO_{it-12} | | | -0.0022 (0.0023) | -0.0021 (0.0022) | -0.0022 (0.0022) | | | -0.0001 (0.0001) | -0.0001 (0.0001) | -0.0001 (0.0001) |
| $ApprovalLength_j$ | -0.0013* (0.0007) | -0.0012* (0.0007) | -0.0016*** (0.0007) | -0.0016*** (0.0007) | -0.0016*** (0.0007) | -0.0008 (0.0005) | -0.0008 (0.0005) | -0.0009* (0.0005) | -0.0009* (0.0005) | -0.0009* (0.0005) |
| M_{ip} | 0.0008 (0.0032) | 0.0005 (0.0032) | 0.0013 (0.0037) | 0.0013 (0.0037) | 0.0014 (0.0037) | -0.0046 (0.0033) | -0.0049 (0.0033) | -0.0057 (0.0040) | -0.0057 (0.0040) | -0.0055 (0.0040) |
| \overline{TO}_{ip} | -0.0035 (0.0080) | -0.0040 (0.0081) | -0.0031 (0.0086) | -0.0030 (0.0086) | -0.0029 (0.0086) | -0.0032 (0.0071) | -0.0037 (0.0072) | -0.0023 (0.0071) | -0.0023 (0.0071) | -0.0020 (0.0070) |
| \overline{NME}_{ip} | 0.4992*** (0.1344) | 0.5099*** (0.1345) | 0.4814*** (0.1339) | 0.4863*** (0.1342) | 0.4777*** (0.1343) | | | | | |
| \overline{NDA}_{ip} | | | | | | 0.6147*** (0.1284) | 0.6255*** (0.1285) | 0.6070*** (0.1269) | 0.6116*** (0.1271) | 0.6029*** (0.1273) |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Table 9 (continued)

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|--------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NME_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} | NDA_{it} |
| Constant | -8.2348*** (0.8982) | -8.4603*** (0.9486) | -8.4407*** (0.9488) | -8.4926*** (0.9424) | -8.3348*** (0.9440) | -7.7385*** (0.8361) | -7.9334*** (0.8486) | -7.9770*** (0.8611) | -8.0133*** (0.8559) | -7.8716*** (0.8529) |
| N | 2169 | 2169 | 1985 | 1985 | 1985 | 2169 | 2169 | 1985 | 1985 | 1985 |
| Indications | 168 | 168 | 163 | 163 | 163 | 168 | 168 | 163 | 163 | 163 |

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses

All market size measures enter in logarithmic form into the analysis

References

- Acemoglu D, Linn J (2004) Market size in innovation: theory and evidence from the pharmaceutical industry. *Quart J Econom* 119(3):1049–1090
- Adner R, Levinthal D (2001) Demand heterogeneity and technology evolution: implications for product and process innovation. *Manag Sci* 47(5):611–628
- Andersen B (1998) The evolution of technological trajectories 1890–1990. *Struct Chang Econ Dyn* 9(1):5–34
- Andersen B (1999) The hunt for s-shaped growth paths in technological innovation: a patent study. *J Evol Econ* 9(4):487–526
- Azoulay P (2004) Capturing knowledge within and across firm boundaries: evidence from clinical development. *Am Econ Rev* 94(5):1591–1612
- Barge-Gil A, López A (2014) R&D determinants: accounting for the differences between research and development. *Res Policy* 43(9):1634–1648
- Becker W, Peters J (2000) Technological opportunities, absorptive capacities, and innovation. Discussion Paper Series, University of Augsburg, Institute for Economics, p 195
- Bitzer J, Stephan A (2007) A Schumpeter-inspired approach to the construction of R&D capital stocks. *Appl Econ* 39(2):179–189
- Blundell R, Griffith R, Reenen JV (1995) Dynamic count data models of technological innovation. *Econ J* 105(429):333–344
- Blundell R, Griffith R, Windmeijer F (2002) Individual effects and dynamics in count data models. *J Econ* 108(1):113–131
- Cameron AC, Trivedi PK (2013) Regression analysis of count data, 2nd edn. Cambridge University Press, Cambridge
- Castellacci F (2007) Technological regimes and sectoral differences in productivity growth. *Ind Corp Chang* 16(6):1105–1145
- Cerda RA (2007) Endogenous innovations in the pharmaceutical industry. *J Evol Econ* 17(4):473–515
- Cockburn IM, Henderson RM (2001) Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. *J Health Econ* 20(6):1033–1057
- Cohen WM, Levin RC (1989) Empirical studies of innovation and market structure. In: Schmalensee R, Willig R (eds) *Handbook of industrial organization*. Elsevier, Amsterdam, pp 1059–1107
- Cohen WM, Levinthal DA (1989) Innovation and learning: the two faces of R&D. *Econ J* 99(397):569–596
- Cohen WM, Levin RC, Mowery DC (1987) Firm size and R&D intensity: a re-examination. *J Ind Econ* 35(4):543–565
- Cohen W, Nelson R, Walsh J (2002) Links and impacts: the influence of public research on industrial R&D. *Manag Sci* 48(1):1–23
- DiMasi JA, Grabowski HG (2007) The cost of biopharmaceutical R&D: is biotech different? *Manag Decis Econ* 28(4-5):469–479
- Dosi G (1988) Sources, procedures, and microeconomic effects of innovation. *J Econ Lit* 26(3):1120–1171
- Fabrizio KR, Thomas LG (2012) The impact of local demand on innovation in a global industry. *Strateg Manag J* 33(1):42–64
- Food and Drug Administration (FDA) (2011) Old NDC Database information. <http://www.fda.gov/Drugs/InformationOnDrugs/ucm257245.htm>, visited on June 18, 2013
- Freeman C (1994) The economics of technical change. *Camb J Econ* 18(5):463–514
- Galambos L, Sturchio JL (1998) Pharmaceutical firms and the transition to biotechnology: a study in strategic innovation. *Bus Hist Rev* 72(2):250–278
- Gambardella A (1995) *Science and innovation: the U.S. pharmaceutical industry during the 1980s*. Cambridge University Press, Cambridge
- Geroski PA, Walters CF (1995) Innovative activity over the business cycle. *Econ J* 105(431):916–928
- Giaccotto C, Santerre RE, Vernon JA (2005) Drug prices and research and development investment behavior in the pharmaceutical industry. *J Law Econ* 48(1):195–214
- Godin B (1996) Research and the practice of publication in industries. *Res Policy* 25(4):587–606
- Gourieroux C, Monfort A, Trognon A (1984) Pseudo maximum likelihood methods: applications to poisson models. *Econometrica* 52(3):701–720
- Grabowski HG (1968) The determinants of industrial research and development: a study of the chemical, drug, and petroleum industries. *J Polit Econ* 76(2):292–306
- Grabowski H, Vernon J (2000) The determinants of pharmaceutical research and development expenditures. *J Evol Econ* 10(1–2):201–215

- Grabowski HG, Wang YR (2006) The quantity and quality of worldwide new drug introductions 1982–2003. *Health Aff* 25(2):452–460
- Grabowski HG, Vernon JM, Thomas LG (1978) Estimating the effects of regulation on innovation: an international comparative analysis of the pharmaceutical industry. *J Law Econ* 21(1):133–163
- Griliches Z, Hall BH, Pakes A (1991) R&D, patents, and market value revisited: is there a second (technological opportunity) factor? *Econ Innov New Technol* 1(3):183–201
- Hamilton KS (2003) Subfield and level classification of journals. CHI Research Inc., CHI No. 2012-R
- Henderson R, Cockburn I (1994) Measuring competence? Exploring firm effects in pharmaceutical research. *Strateg Manag J* 15:63–84
- Jaffe AB (1986) Technological opportunity and spillovers of R&D: evidence from firms' patents, profits, and market value. *Am Econ Rev* 76(5):984–1001
- Jaffe AB (1989a) Characterizing the 'technological position' of firms, with application to quantifying technological opportunity and research spillovers. *Res Policy* 18(2):8–97
- Jaffe AB (1989b) Real effects of academic research. *Am Econ Rev* 79(5):957–970
- Kleinknecht A, Verspagen B (1990) Demand and innovation: Schmookler re-examined. *Res Policy* 19(4):387–394
- Klevorick AK, Levin RC, Nelson RR, Winter SG (1995) On the sources and significance of interindustry differences in technological opportunities. *Res Policy* 24(2):185–205
- Kneller R (2010) The importance of new companies for drug discovery: origins of a decade of new drugs. *Nat Rev Drug Discov* 9(11):867–882
- Kremer M (2002) Pharmaceuticals and the developing World. *J Econ Perspect* 16(4):67–90
- Lacy C, Armstrong L, Goldman M, Lance L (eds) (2010) *Drug information handbook*, 19th edn. Lexi-Comp, Hudson
- Levin RC, Cohen WM, Mowery DC (1985) R&D appropriability, opportunity, and market structure: new evidence on some Schumpeterian hypotheses. *Am Econ Rev* 75(2):20–24
- Levin RC, Klevorick AK, Nelson RR, Winter SG, Gilbert R, Griliches Z (1987) Appropriating the returns from industrial research and development. *Brook Pap Econ Act* 1987(3):783–831
- Lichtenberg FR (2004) Sources of U.S. longevity increase, 1960–2001. *Q Rev Econ Fin* 44(3):369–389
- Lichtenberg FR (2005a) Pharmaceutical innovation and the burden of disease in developing and developed countries. *J Med Philos* 30(6):663–690
- Lichtenberg FR (2005b) The impact of new drug launches on longevity: evidence from longitudinal, disease-level data from 52 countries, 1982–2001. *Int J Health Care Fin Econ* 5(1):47–73
- Lim K (2004) The relationship between research and innovation in the semiconductor and pharmaceutical industries (1981–1997). *Res Policy* 33(2):287–321
- Mahlich JC, Roediger-Schluga T (2006) The determinants of pharmaceutical R&D expenditures: evidence from Japan. *Rev Ind Organ* 28(2):145–164
- Malerba F, Nelson R, Orsenigo L, Winter S (2007) Demand, innovation, and the dynamics of market structure: the role of experimental users and diverse preferences. *J Evol Econ* 17(4):371–399
- Mansfield E (1991) Academic research and industrial innovation. *Res Policy* 20(1):1–12
- McKelvey M (1996) *Evolutionary innovations: the business of biotechnology*. Oxford University Press, Oxford
- Morton FMS (1999) Entry decisions in the generic pharmaceutical industry. *RAND J Econ* 30(3):421–440
- Mowery D, Rosenberg N (1979) The influence of market demand upon innovation: a critical review of some recent empirical studies. *Res Policy* 8(2):102–153
- Narin F, Olivastro D (1992) Status report: linkage between technology and science. *Res Policy* 21(3):237–249
- Narin F, Rozek RP (1988) Bibliometric analysis of U.S. pharmaceutical industry research performance. *Res Policy* 17(3):139–154
- Nelson RR (1988) Modelling the connections in the cross section between technical progress and R&D intensity. *RAND J Econ* 19(3):478–485
- Nelson RR, Wolff EN (1997) Factors behind cross-industry differences in technical progress. *Struct Chang Econ Dyn* 8(2):205–220
- Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics* 57(1):120–125
- Rasmussen B (2010) *Innovation and commercialisation in the biopharmaceutical industry: creating and capturing value*. Edward Elgar Publishing Ltd, Cheltenham and Northampton
- Reiffen D, Ward MR (2005) Generic drug industry dynamics. *Rev Econ Stat* 87(1):37–49

- Scherer FM (1965) Firm size, market structure, opportunity, and the output of patented inventions. *Am Econ Rev* 55(5):1097–1125
- Scherer FM (1982) Demand-pull and technological invention: Schmookler revisited. *J Ind Econ* 30(3):225–237
- Schmookler J (1966) *Invention and economic growth*. Harvard University Press, Cambridge
- Toole A (2007) Does public scientific research complement private investment in research and development in the pharmaceutical industry? *J Law Econ* 50(1):81–104
- Toole AA (2012) The impact of public basic research on industrial innovation: evidence from the pharmaceutical industry. *Res Policy* 41(1):1–12
- von Hippel E (1976) The dominant role of users in the scientific instrument innovation process. *Res Policy* 5(3):212–239
- Walsh V (1984) Invention and innovation in the chemical industry: demand-pull or discovery-push? *Res Policy* 13(4):211–234
- Yin W (2008) Market incentives and pharmaceutical innovation. *J Health Econ* 27(4):1060–1077