KU LEUVEN
FACULTY OF ECONOMICS AND BUSINESS
Department of Management, Strategy and Innovation

Prescribing Science

How does science influence innovation in pharmaceutical firms?



Dissertation presented
to obtain Doctor in
Business economics
by
Relinde Colen

Number 617 2018

Daar de proefschriften in de reeks van de Faculteit Economie en Bedrijfswetenschappen het persoonlijk werk zijn van hun auteurs, zijn alleen deze laatsten daarvoor verantwoordelijk.

Doctoral Committee

Advisor

Prof. Dr. René Belderbos	KU Leuven, Maastricht University, UNIMERIT

Co-advisors

Prof. Dr. Stijn Kelchtermans	KU Leuven, U Hasselt
Prof. Dr. Bart Leten	KU Leuven, U Hasselt

Members

Prof. Dr. Hanna Hottenrott	TUM
Prof. Dr. Ammon Salter	University of Bath
Prof. Dr. Reinhilde Veugelers	KU Leuven, Breugel, CEPR

Chairperson

Prof. DR. Marc Lambrecht	KU Leuven

Acknowledgements

During my master year, I was asked to start a Ph.D. However, I confidently replied 'No'. I had enough of the university, of reading, of theory. I wanted a hands-on job where I would have immediate results. Hence, management-trainee at Colruyt n.v. was the job I took on after graduation. In this position and thanks to my colleagues, I grew a lot as a person and a leader. However, I kept asking myself that one question: "Would you like to start a Ph.D.?". In 2013, I then found my answer and started my Ph.D. at the MSI department. To my great pleasure, I can say that I have no regrets and that this whole process was the one I needed to take to reach this stage. I have enjoyed my Ph.D. and will continue within academic as, for me, it provides the right balance between (1) challenges, freedom and independence, and (2) peer recognition, interaction and debate. While academics do not have a direct impact, I hope that my teaching can trigger the critical mind of students and that my work will contribute in the joint development of new insights. This being said, the pursued of a Ph.D. had a large impact on my life. Fortunately, I have always been surrounded by highly encouraging and helpful people.

First, I would like to show my greatest appreciation to my supervisors. Not only, did they provide a good starting point so that I could attend conferences and meet the scientific community already in my first year. They have also given me detailed and highly-informed feedback during my whole trajectory. Our meetings were always a pleasure to attend with vivid debates and frequent laughs. Moreover, they have always been available for a talk or a 'quick' email response. René, thank you for guiding me so consistently and professionally. With your experience and commitment, you showed me the tricks of the trade to hopefully become a successful academic. I would like to thank Stijn for his high understanding of my workload and frustrations. Thanks to his broad knowledge and openness to debate, he has pushed me to develop my own opinions and scientific style. Finally, I would like to show my gratitude to Bart who was not only the person who asked me to start a Ph.D. but also frequently asked how I was doing during the process. Thank you for those short coffee breaks which often leaded to revived enthusiasm and new reflections. While I also appreciate all the feedback that you gave me, I am mostly thankful for the times you came to me for advice showing me that you appreciate my opinion and recognize my scientific skills.

My gratitude also goes to the other members of my Ph.D. committee: Reinhilde Veugelers, Hanna Hottenrott and Ammon Salter. Your insightful comments will greatly improve my work.

Not only, did and will they have a positive influence on the studies of this dissertation, they also left a long lasting impression on me as a scientist. Reinhilde Veugelers, you taught me to not think about the question to be asked but about the answer that can be found. While curious scientists see interesting questions everywhere, they need to reflect and select that question of which the answer that can be formulated is of the highest value. Moreover, you, among others, inspired me to remain in academia by the enthusiasm and joy you show as a scientist. Hanna Hottenrott, I would like to additionally thank you for all the comments and advice you gave me at conferences and during my visit at TUM. I really enjoyed our talks in München and will take your advice to hard. You, for instance, showed me the strength of descriptive statistics in both idea generation and argumentation. Ammon Salter, thank you for your experienced and strong comments. Within science, it is important to have a critical look at your work and have the strength to adjust your ideas and expectations.

I also had the pleasure to meet many amazing people at our department and during conferences. They not only gave useful comments on my work, they have also embraced me in the unique scientific community. I would like to express my gratitude to Pablo D'este and Massimo Riccaboni who welcomed me at their institution during a short stay. Pablo, thank you for taking the time to show me around and closely involving me in the research group. Additionally, I would like to thank Maikel and Antonio for showing interest in my work from the start and sharing their realistic view on a Ph.D. dissertation and post-doc. Rossella and Dennis, by sharing your experiences as a Ph.D. student and thoughts on an academic career, you spurred my enthusiasm for scientific research and helped me in making the decision to remain in academia. I would also like to show my gratitude to my office mates who were always open for a talk or there with a helping hand. During my life, I have realized the importance of colleagues and I must say that I have been blessed during my Ph.D. In general, I would like to thank all members of the department for the good atmosphere I have worked in.

Also outside the scientific community, some deserve a special word of appreciation. Most importantly, I would like to thank my husband who supported my career change and endured my busy schedule and stressful moments. A special thanks for your realistic reflections and the moments where you and our son reminded me what really counts in live. Also, my parents deserve a loud "Thank you". At every step in my life, they have given me the freedom and support I needed. They have not only showed me what it means to work hard but also how to enjoy, especially the small things, in life. I would also like to express my gratitude to my parents in law for always being there when we needed help. Moreover, I would like to thank Hanne for

the long energizing talks on the train and for watching Matthijs. My brother Pieter, also played an important role during my Ph.D., not only has he been an example for me. He has also often taken the time to proofread my work and to reassure me at times of demotivation. I also would like to extend a word of thanks to Katlijn, a good friend who proceeded me on the doctoral journey. Her stories on academic life have, among others, triggered me to start a Ph.D. and pursue an academic career. Finally, I am grateful for my whole family and my friends who have always followed my Ph.D. with interest and showed their strong believe in my capabilities.

Relinde Colen

Leuven, September 2018

Table of Contents

Doctoral Committee ii
Acknowledgements iii
Table of Contents vi
List of Tables viii
List of Figures x
General Introduction
Science intensive industry 1
Productivity crisis 2
Prescribing science: overview of the three chapters 5
Data 9
Chapter 1: Scientific Research, Drug Candidates and Drug Development
Abstract 15
1.1 Introduction 16
1.2 Importance of science for innovation 18
1.3 The Productivity Crisis in the Pharmaceutical Industry 20
1.4 Scientific Background of the Drug Candidate 22
1.5 The Market for Technology in the Pharmaceutical Industry 24
1.6 Data and Descriptive Statistics 27
1.7 Empirical Results 35
1.8 Supplementary Analyses 39
1.9 Discussion 45
1.10 Conclusion 47
1.11 appendix 50
Chapter 2: Reaching for the Stars
Abstract 63
2.2 theoretical background and hypotheses 68
2.3 data, variables and methods 79
2.4 empirical results 90
2.5 supplementary analysis 95
2.6 conclusion and discussion 98
2.7 Appendix 102

Chapter 3: Firm entry into new scientific areas
Abstract 115
3.1 Introduction 116
3.2 Background theory 120
3.3 Connections by insiders and outsiders 124
3.4 Hypotheses 126
3.5 Data 129
3.6 Variables and descriptive statistics 134
3.7 Regression results 141
3.8 Supplementary analysis 146
3.9 Discussion 150
3.10 Conclusion 152
3.11 Appendix 156
General Conclusion
Science and the productivity crisis 169
Implications 171
Limitations 172
Future research directions 173
References 175

List of Tables

T**able 1.** Overview of the three chapters 6
Table 1.1. Descriptive statistics on SNPRs within the total set and subgroups of taken-up of drug candidates 30
Table 1.2. Descriptive statistics on the control variables within the total set and subgroups of taken-up drug candidates 32
Table 1.3. Correlation matrix of the regression variables 34
Table 1.4. Logit regression on the probability of take-up into clinical trials: Coefficients 36
Table 1.5. Multinomial Logit regression results explaining drug candidates not being taken up (1), taken up by other firms (2) and taken up by patent owner (3): Coefficients 37
Table 1.A1. Logit regression on the probability of take-up into clinical trials: A.M.E. 50
Table 1.A2. Multinomial Logit regression results explaining drug candidates not being taken up (1), taken up by other firms (2) and taken up by patent owner (3): A.M.E. 51
Table 1.A3. Logit and Multinomial Logit regression results. Take-up defined as the start of pre-clinical research 52
Table 1.A4. Logit regressions explaining development success (pre-registration), by other firm and by patent owner 53
Table 1.A5. Logit regressions explaining development success, measured by registration (rather than pre-registration), by other firm and by patent owner 54
Table 1.A6. Logit and Multinomial Logit regression results, controlling for only external SNPRs 55
Table 1.A7. Logit and Multinomial Logit regression results, controlling for stated discovery and preclinical research fitting the firm's scientific strategy 56
Table 1.A8. Multinomial Logit regression results for successful firms 57
Table 1.A9. Multinomial Logit regression results for unsuccessful firms 58
Table 1.A10. Multinomial Logit regression results for take-up limited to patent applications at USPTO 59
Table 1.A11. Logit and Multinomial Logit regression results for take-up restricting sample to patents in IPC class A61P 60
Table 1.A12. Logit and Multinomial Logit regression results for take-up restricting sample to patents in IPC class A61P and (C07 or C12) 61
Table 1.A13. Logit regression results explaining whether drug candidates are taken up by other firms or by the patent owner 62
Table 2.1. Descriptive statistics (406 obs.) 86
Table 2.2. Correlation matrix of all variables (406 obs.) 89
Table 2.3. Academic star collaboration in basic research and firms' innovation performance: Pseudo fixed effects Poisson panel regression results 92
Table 2.4. Tests for the hypotheses on innovation performance benefits 94
Table 2.A1. Academic star collaboration in basic research and firms' innovation performance: firm-star matching 102
Table 2.A2. Academic star collaboration in basic research and firms' innovation performance: collaborative research quality heterogeneity 104
Table 2.A3. Academic star collaboration in basic research and firms' innovation performance: collaborative research experience heterogeneity 106

Table 2.A4. Academic star collaboration in basic research and firms' innovation performance: 8-year citation window for the dependent variable 108
Table 2.A5. Reverse causality? Firms' prior patent performance and engagement in different collaboration types 110
Table 2.A6. Investigating star heterogeneity: description and descriptive statistics of the variables 111
Table 2.A7. Investigating star heterogeneity: Probit regression results robust for heteroscedasticity 112
Table 2.A8. Firms included in the sample 113
Table 3.1. Descriptive statistics of variables of interest across all observation, observations of no-entry and observations of entry. 132
Table 3.2. Correlation matrix of the regression variables 140
Table 3.3. Cloglog panel regression with random effects predicting firm's probability to enter a research area 142
Table 3.4. Cloglog panel regression with random effects predicting firm's probability to enter a research area, examining the social connections to scientists in the area 144
Table 3.5. Descriptive statistics on the marginal effects of the variables of interest considering the other variables as observed within the sample 145
Table 3.A1. Cloglog regression predicting firm's probability to enter a research area persistently (at least 5 publication in coming 8 years) 156
Table 3.A2. Descriptive statistics on firms' publications in the research area in the first two years since entry 157
Table 3.A3. Two-stage Poisson regression predicting firm's publication count in the first two years of entering the research area 158
Table 3.A4. Cloglog panel regression with random effects predicting firm's probability to enter a research area controlling for first-stage strength of the connections 159
Table 3.A5. Cloglog panel regression with random effects predicting firm's probability to enter a research area controlling for stardom in first line 161
Table 3.A6. Cloglog random effects regression predicting firms' probability to enter a research area without firm dummies 163
Table 3.A7. Cloglog panel regression with random effects predicting firm's probability to enter a research area, split sample according to the scientific absorptive capacity of the firm 165
Table 3.A8. Cloglog random effects regression predicting firms' probability to enter a research area explained by social connections established from t-6 to t-3 167

List of Figures

Figure 1.1. The drug development process..... 21
Figure 2.1. Construction and descriptive statistics of the key collaboration variables..... 84
Figure 2.2. the key variables of basic research collaboration across observed innovation performance 85
Figure 3.1. Firm's social network ties to a new scientific research area "A"..... 125
Figure 3.2. The formation of connections to the area across the years before entry 136
Figure 3.3. The formation of insider connections to area's scientists across the years before entry.. 136

General Introduction

The pharmaceutical industry is a recurrent subject of analysis within the innovation literature. Not only is the industry empirically interesting for its large data availability, it is also of enormous social importance both for economic growth (Kola and Landis 2004; Lichtenberg 2005) as for health improvement (Scherer, 2010). Furthermore, its extraordinary investments in fundamental research and strong university-industry links have made it a textbook example of a science-intensive industry. However, since the nineties concerns have been raised on the research productivity of bio-pharmaceutical companies. While their R&D expenditures are increasing steadily, the number of new products on the market is dropping. Both practitioners and innovation scholars are speculating on the reasons behind this so-called 'productivity crisis' referring to the increasing challenge of scientific discovery (Danzon, Nicholson, and Pereira 2005; Crispeels, Willems, and Scheerlinck 2018; Moors, Cohen, and Schellekens 2014; Pammolli, Magazzini, and Riccaboni 2011), the lack of fundamental and translational understanding (Kola and Landis 2004; I. Cockburn 2007) and the growing organizational barriers between research and development (Arora et al. 2009; I. Cockburn 2007).

By analyzing different performance criteria across the drug development process, this Ph.D. dissertation hopes to expand our understanding on the role of science within the pharmaceutical industry. Moreover, reflecting on the productivity crisis, I aim to provide some recommendations on the management and organization of science that can assist firms in enhancing the relationship between their R&D expenditures and performance.

In the following paragraphs, I will first describe the extra-ordinary presence of science within the pharmaceutical industry. Following, I will discuss the different opinions formulated on the causes of the productivity crisis mentioned in the literature. Moreover, I will provide some information on the data analyzed in his dissertation. Finally, I will position the three papers of the dissertation within the context of the productivity crisis.

SCIENCE INTENSIVE INDUSTRY

Multiple studies have indicated that pharmaceutical products are highly dependent on scientific discoveries (Cockburn & Henderson, n.d.; Maxwell & Eckhardt, 1990, McMillan *et al.*, 2000; Collins and Wyatt, 1998). Mansfield (1998), for instance, asked laboratory heads of 77 firms across 7 broad technology fields which percentage of their innovations between 1975 and 1985

(1986-1994) could not have been developed in the absence of recent academic research. While the lab heads reported an average research dependence of 10 % (11%) across products, they reported a dependence of 27 % (31%) for drugs and medical products. Moreover, in comparison to other fields, patents in drugs and medicine cite significantly more scientific publications, especially within basic research journals (Narin, Hamilton, and Olivastro 1997). Additionally, pharmaceutical companies are significantly larger investors in research and development than manufacturing companies in the US (Scherer 2000). They are even known as strong investors of basic-research projects, aimed at the enhancement of fundamental understanding without a predetermined application. In 1993, drug companies devoted 13.6 % of their R&D budget on basic research while other research-performing companies only assigned 6.3 % (Scherer 2000).

Multiple studies have shown that these internal scientific activities are related to higher innovation performance, measured by (citation weighted) patent counts (Arora, Belenzon, & Patacconi, 2015; Gambardella, 1992; Kelchtermans, Leten, & Belderbos, 2013), average patent quality (Fabrizio 2009), number of new molecular entities, radical innovation (Della Malva *et al.*, 2015) and stock market value (Arora, Belenzon, and Patacconi 2015). Although firms' profit-seeking and application-oriented culture is not well aligned with the long-term and uncertain nature of basic research, scholars have emphasized that there are benefits for performing basic research in-house: (1) enhanced understanding of the technological landscape facilitating the search process to new inventions and steering applied research efforts, and (2) increased ability to absorb external research (Rosenberg 1990; Stijn Kelchtermans, Leten, and Belderbos 2013; Arora, Belenzon, and Patacconi 2015; Arora, Belenzon, and Patacconi 2017). Furthermore, studies have indicated that firms' internal scientific activities are more strongly related to innovation performance when involving academic collaboration partners (Almeida, Hohberger, and Parada 2011; Cockburn and Henderson 1998; Fabrizio 2009; Belderbos, Carree, and Lokshin 2004; Belderbos, Gilsing, and Suzuki 2016a).

PRODUCTIVITY CRISIS

Since the nineties, innovation scholars are repeatedly providing evidence on the important role of science within the pharmaceutical industry, though, companies are being confronted with a diminishing return on their research and development investments (I. Cockburn 2007; Kola and Landis 2004). While there was a clear drop in productivity from the nineties to 2005, the situation appeared to stagnate in 2005 (Pammolli, Magazzini, and Riccaboni 2011) and even

recover in 2010. However, recent figures are again showing dropping returns in productivity (Colin and Taylor 2018). Scholars have been speculating on possible causes of this 'productivity crisis': increasing challenge of scientific discovery (Danzon, Nicholson, and Pereira 2005; Crispeels, Willems, and Scheerlinck 2018; Moors, Cohen, and Schellekens 2014; Pammolli, Magazzini, and Riccaboni 2011), the lack of basic and translational understanding (e.g. Kola and Landis, 2004; Cockburn, 2007) and the growing organizational barriers between research and development (Arora et al. 2009; I. Cockburn 2007).

Increasing challenge of scientific discovery

Some scholars argue that the decreasing ratio of new drugs on the market per invested dollar in R&D is merely an indication that the low hanging fruit has been picked (I. Cockburn 2007; Sams-Dodd 2005; Pammolli, Magazzini, and Riccaboni 2011). They argue that the 'easy' health problems have been solved and that to capture the remaining research opportunities firms require to solve highly complex problems and make considerable investments in new research capabilities.

Pharmaceutical firms have been under pressure to follow novel and distant research opportunities because the emergence of reference pricing schemes, the increased price competition and the raise of generic alternatives has decreased the profitability of incremental innovation and follow-on drugs (Pammolli, Magazzini and Riccaboni, 2011). The most discussed opportunity that emerged during the period of downturn is biotechnology which changed the drug discovery approach from physiological-based to target-based (Sams-Dodd 2013). Additionally, also other new tools and methods appeared like throughput screening, combinatorial chemistry, micro fluids and bioinformatics (I. Cockburn 2007). Moreover, thanks to extraordinary advancements in basic science, opportunities emerged in new disease areas like Alzheimer's. While the progress in technologies and research area's has led to a considerable increase of 'druggable targets', substantial investments have been necessary to acquire the specialized assets, human capital and organizational infrastructure necessary for their discovery and development (I. Cockburn 2007). Moreover, spurred by globalization knowledge has increasingly dispersed across geographical hotspots further enlarging the struggle to remain informed on all development (Singh 2008; Lahiri 2010). Additional to significant costs, the new research opportunities are also related to higher risks as they focus on more complex diseases on which there is lower fundamental understanding and development experience (Pammolli, Magazzini, and Riccaboni 2011; Iain M. Cockburn 2006).

Lack of basic and translational understanding

The lack of fundamental understanding and translational research has been put forward as a root cause of the productivity crisis by many scholars (Iain M. Cockburn 2006; Sams-Dodd 2013), especially within the life sciences themselves (Kola and Landis 2004; Waring et al. 2015; Barrenho, Miraldo, and Smith 2013; Cook et al. 2014; J.-Y. Chiou et al. 2016; Hay et al. 2014; Sams-Dodd 2013). Moreover, it forms a major source of critique on the target-based drug discovery approach that accompanied biotechnology (Sams-Dodd 2013). While pharmaceutical firms traditionally discovered new drugs without knowing the biological basis of the disease or the mechanism by which the compound exerted her effects, this new approach demands for a fundamental understanding of the disease to identify possible molecules that may form effective treatments (Sams-Dodd 2013). However, for the highly complex and novel health problems, which firms especially aim to tackle since the nineties, there is not only limited fundamental understanding, but also deficient verification of supporting technologies and methods, like animal models (Sams-Dodd 2013).

While this situation makes some scholars criticize the use of target-based discovery (Sams-Dodd 2005; Sams-Dodd 2013), others emphasize the necessity of stronger investments in basic and translational research (Schuhmacher *et al.* 2016, p 123; Kola and Landis, 2004; Barrenho, Miraldo and Smith, 2013; Cook *et al.*, 2014; Hay *et al.*, 2014; Waring *et al.*, 2015; Chiou *et al.*, 2016). The latter have observed that drug candidates often suffer from mismatches between disease, compound, target group and clinical techniques which could have been foreseen and even resolved with a better fundamental understanding of human physiology or disease characteristics, and a stronger translational skillset to select adequate animal models and drug transportation mechanisms (Kola and Landis 2004; Waring et al. 2015; Barrenho, Miraldo, and Smith 2013; Cook et al. 2014; J.-Y. Chiou et al. 2016; Hay et al. 2014). For example, Paul and colleagues (2010) identified the doubtful validity of the target for the disease pathophysiology as a key reason for dropout of a drug candidate, besides inadequate compound characteristics or safety issues (Eichler et al. 2009). The fact that the high failure rates in drug development are mostly based on a lack of efficacy – whether the clinical trials are in line with the promising lab results – supports this concern of limited basic and translational understanding (Kola and Landis 2004; Waring et al. 2015; Barrenho, Miraldo, and Smith 2013; Cook et al. 2014; J.-Y. Chiou et al. 2016; Hay et al. 2014).

Organizational barriers between research and development

Within the innovation literature more attention has been given to the structural change of the pharmaceutical sector during the eighties and nineties (Cockburn, 2004; Cockburn, 2007; Stuart, Ozdemir and Ding, 2007; Arora *et al.*, 2009; Malik, 2011; Belderbos, Gilsing and Suzuki, 2016). While traditionally, pharmaceutical companies were vertically integrated firms active in both research and development, the value chain has disaggregated with some players focusing on research, others on development and a few remaining highly integrated. Complementary, an active market for technology has been established with increasing vertical transactions like licensing of drugs and drug candidates, joint ventures and strategic technology alliances, and mergers and acquisitions (Malik 2011). As scientific research and product development are two distinct tasks with their own needs for specific resources and capabilities, the structural change could stimulate productivity as it divides tasks across more specialized actors. In support of this reasoning, Chesbrough and Chen (2013) emphasize the importance of licensing out drug candidates that can turn out to be successful within other firms. However, questions have been raised on the effectiveness of this changed structure within a science-intensive industry like pharmaceuticals (I. Cockburn 2007; Arora and Gambardella 1994).

Some scholars have indicated that disaggregation and division of labor is only optimal in perfect markets with complete information, minimal transaction costs and fluid transfer of assets (Iian M. Cockburn 2004; I. Cockburn 2007; Arora and Gambardella 1990). In contrast, pharmaceutical R&D is characterized by long-term, risky projects entailing complex and sticky knowledge, difficult to monitor activities, substantial project-specific unrecoverable investments and vertically complementary outcomes (I. Cockburn 2007). As under these circumstances licensing entails numerous difficulties (Pisano, 1997; Chesbrough and Chen, 2013), some scholars rate disaggregation and usage of the market for technology inferior to vertical integration and see it as a possible cause of the decreasing productivity in the pharmaceutical industry (Arora and Gambardella, 1994; Cockburn, 2004; Cockburn, 2007).

PRESCRIBING SCIENCE: OVERVIEW OF THE THREE CHAPTERS

When confronted with the productivity crisis, pharmaceutical companies may turn to innovation scholars and their findings for salvation. As indicated above, prior literature has signaled that investments in in-house scientific (basic) research and collaboration with academia are regarded as triggers for innovation performance within the pharmaceutical

industry (Rosenberg, 1990; Cockburn and Henderson, 1998; Belderbos, Carree and Lokshin, 2004; Fabrizio, 2009; Almeida, Hohberger and Parada, 2011; Kelchtermans, Leten and Belderbos, 2013; Arora, Belenzon and Patacconi, 2015, 2017; Belderbos, Gilsing and Suzuki, 2016). However, it remains unclear whether this prescription of science would help firms in recovering their productivity. More precisely, practitioners may benefit from knowing which scientific efforts would form an effective treatment for particular 'causes' of the productivity decline as not every firm may suffer from the same difficulties.

With this dissertation, I hope to contribute to fulfilling this challenge. I investigate three different performance criteria across the drug-development process acknowledging the broad benefits assigned to scientific insight within the pharmaceutical industry: 1) development of drug candidates, 2) patent performance and 3) entry into new scientific areas. The different settings of the three chapters additionally allow for a discussion on the benefit of particular scientific efforts in response to the suggested causes of productivity decline. Table 1 presents a schematic overview of the three chapters with investigated performance criteria and related causes of productivity decline. In the following, the three chapters are briefly described.

Table 1. Overview of the three chapters

	Chapter 1	**Chapter 2**	**Chapter 3**
Investigated performance criteria	Development of drug candidates (internally or externally)	Patenting performance (citation weighted patent count)	Entry into new disease areas
Investigated scientific efforts	In-house basic and applied research	Basic-research collaboration with academic star scientists	Development of a social network within scientific communities through collaboration
Related causes of productivity decline	Lack of basic and translational understanding Organizational barriers between research and development	Increasing challenge of capturing opportunities Lack of basic and translational understanding	Increasing challenge of capturing opportunities

Chapter 1: Scientific Research, Drug Candidates and Drug Development

The literature indicates that, even during the productivity crisis, an abundance of new drug candidates are invented yearly (I. Cockburn 2007). However, as the development of these candidates entails substantial and increasing costs at highly uncertain success rates (DiMasi, Hansen, and Grabowski 2003; DiMasi, Grabowski, and Hansen 2016), firms are highly diligent on starting up the development process internally or allowing other firms to do so. On average, for only 5 percent of the drug candidates development is initiated, and for 1 percent it is successfully finalized with drug approval and market launch (Campbell, 2005; PPD, 2017). It is important to obtain a better understanding of this forceful decision within the development process. From a societal perspective, it is of importance that inventions, even with low commercial value, are further explored in clinical trials hopefully curing or treating multiple humans in the future. From a company perspective, it is useful to provide indicators of drug candidates that may reflect their potential on take-up and understand what triggers the decision of further development. In this chapter, I investigate whether the development probability of drug candidates is higher when the candidate can rely on some scientific understanding.

The results support the fact that increased scientific understanding, both fundamental (basic) and translational (applied), are valued by firms who aim to develop drug candidates. Moreover, the findings reflect that drug candidates on which the firm has in-house basic background knowledge are more likely to be developed internally. This finding may imply that internal projects may relate to irrational considerations of over-commitment, over-optimism and not-sold-here syndrome leading to a higher likelihood of extended investments from research into development. However, the results show stronger support for the interpretation that the focal firm, as embedder of related fundamental knowledge, may have a competitive advantage in developing the drug candidate. This reasoning implies that there is complementarity between research and development acknowledging the importance of fundamental scientific understanding for development and the difficulty of knowledge transfer across organizational boundaries.

Chapter 2: Reaching for the stars

As a lack of scientific understanding and increasingly complex health problems are considered key concerns within the pharmaceutical industry, collaboration with academic star scientists, who contribute extraordinarily to their research area, may form an attractive strategy to rapidly obtain high-quality capabilities. This chapter investigates whether collaboration with academic star scientists is more strongly related to firms overall ability to produce a higher quantity and

quality of patented inventions than collaboration with average scientists. Patents protect inventions that immediately result from firms' internal or open research activities without being biased by take-up concerns or development issues. Consequently, they are a suited output measurement to capture both the direct innovative result of joint-research as the indirect result of enhanced research capabilities thanks to interaction with academic stars.

The results of the firm-level analysis suggests that on average star scientists are less beneficial collaboration partners for firms aiming to increase their overall patenting performance. However, more detailed examination uncovers important contingencies for beneficial collaboration with academic stars. First, joint basic research with star scientists who collaborate with a single industrial collaboration partner appears to have a significantly stronger relationship with firms' patenting performance. We posit that this 'dedication' of the academic star reflects her commitment of time and effort to transfer her knowledge to the firm during frequent meetings and the protection against possible unintended knowledge spillovers to other industrial collaboration partners. Second, collaboration with academic stars is more strongly related to firms' innovation performance when the academic star scientist also performs applied research with the firm and no internal star scientist is involved or, alternatively, when an internal star scientist is involved and the academic star scientist does not perform applied research with the firm. This finding signals the importance of translational research capabilities within the collaborative team while warning for possible conflict or redundancy when assigning multiple stars to the same team.

Chapter 3: Firm entry into new scientific areas

While the prior literature and the past two chapters have indicated the importance of scientific activities at pharmaceutical firms, it remains unclear how firms determine their portfolio of scientific activities and whether this is influenced by their collaborations within the scientific community. Moreover, the increasing challenge of scientific discovery, explained above, indicates the difficulty of performing scientific research and catching up with scientific progress. Hence, the last chapter of my dissertation investigates firms' entry behavior into new scientific areas. More precise, the chapter examines whether firms' social connections to particular scientific areas may increase firms' likelihood to start up scientific activity in the area. By collaborating with academia, firms may obtain direct and indirect connections to scientists active within a research area. I posit that these connections may provide firms with relevant information on the area assisting them in keeping abreast with the latest advancements and in recognizing valuable research opportunities for the firm. In this chapter, I take into

account all social connections to scientists within the community of a scientific area as all of them have the potential to provide the firm with relevant information. In a supplementary analysis, I do explore the role of star scientists within these social connections.

The results show that social connections to an area only increases firms' probability on entry when the information is communicated to the firm by scientists active in the particular research area. I argue that scientists not active in the area may not receive area relevant information, may not recognize its value, may not be acknowledged as a reliable information source by the firm and may even opportunistically demotivate firms to enter. Additionally, the results show that scientists active in the area can both communicate the knowledge and information she has herself as also effectively exchange the information of her area contacts to the firm.

DATA

These three chapters build upon a worldwide panel dataset of the largest R&D spenders within the biopharmaceutical industry between 1995 and 2002 containing consolidated patent, publication and company information. For analyzing the probability on development in Chapter 1, this dataset is additionally linked to project information of the Pharmaceutical Industry Database (PHID) maintained at the IMT Institute for Advanced Studies in Lucca (Italy). During my Ph.D. trajectory, I have mainly examined and refined the publication data that links information from PubMed (author disambiguation, classification of medical subjects) to affiliation data from Web of Science. I further refined this dataset with complementary data: (1) the Authority dataset (Torvik and Smalheiser 2009) to enable complete author disambiguation, the identification of star scientists and the composition of social networks, (2) self-generated string algorithm for affiliation data to identify university-industry collaborations, (3) the CHI-journal classification to distinguish between basic and applied research.

Co-authorship as approximation of collaboration

Within this dissertation, co-authorship is used as an approximation of collaboration and, hence, social ties. Even though interactions may also occur through other channels, prior research has validated co-authorship as a reliable indicator of collaborative research (e.g. Cockburn and Henderson 1998; Fabrizio 2009). In the study of Melin and Persson (1996), only five percent of the surveyed scientists reported instances of collaboration not resulting in co-authored papers. Additionally, a survey of European and Japanese firms in the electronics and pharmaceuticals sectors reported that the largest majority (84-93%) of co-publications involved

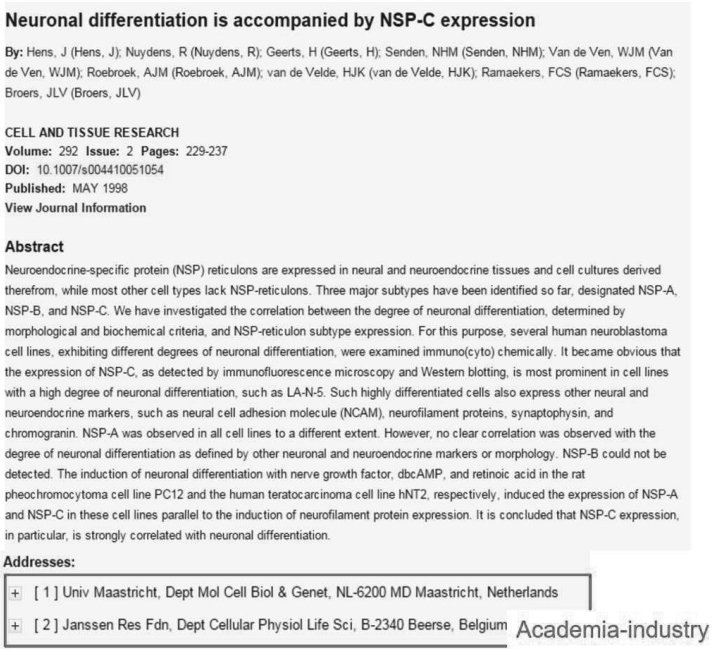
at least some sort of collaboration (Hicks, 1996). In sum, most collaborations result in co-authored publications, and most co-publications do reflect actual research collaborations. Moreover, within the pharmaceutical industry research publications have an important signaling power for firms stimulating them to publish and list their scientists as authors. As the pharmaceutical market is highly regulated and highly dependent on scientific evidence it is crucial to obtain a reputation of scientific excellence to attract R&D employees and partners. Finally, within the second chapter where the focus lies on collaboration between firms and academic stars, we purposely limit ourselves to basic research collaborations which entail close interaction with bidirectional transfer of complex and tacit knowledge. Hence, the research efforts on both sides may likely be rewarded by authorship. Laudel (2002) also indicates that co-publications are a suited measurement for collaborations under joint formulation of the research problem as well as joint realization of research activities.

While co-authorship on a publication is used to signal joint-research, occurrences of co-authorship within the portfolio of firms and scientists is also used to determine their social ties and fabricate social networks. However, to extend this process across publications it is crucial to correctly assign publications to individual authors which can be a daunting task considering the large ambiguity of author names. First, several individuals may be using the same name, like John Smith. Second, the name of individuals may change over the years. For example, women may choose or are obliged to take the sire name of their husband after marriage. Moreover, names often exists out of multiple components of which not all parts are reported consistently. This is especially an issue for nationalities who typically carry multiple sire names like Spaniards. Finally, writing mistakes may occur on the author list of some publications. Fortunately, we were able to utilize the Authority dataset of Torvik and colleagues (Torvik *et al.*, 2005; Torvik and Smalheiser, 2009), who have uniquely identified authors on PubMed publications until 2009.

String algorithm for affiliation data

Figure 1 presents an example publication originating from my dataset entitled "Neuronal differentiation is accompanied by NSP-C expression.". The addresses mentioned on publications provides information on the affiliations where the authors realized most of their work related to the published research. In lines of this dissertation, these addresses were compared with key words of the mother and (majority) daughter companies of the sample firms from 1995 to 2002 resulting in a consolidated set of publications for each pharma and biotech firm. For example, the publication of Figure 1 was assigned to Johnson and Johnson.

Moreover, for each of the corporate publications I investigated whether also an academic institution was involved which would imply that the co-publication results from an university-industry collaboration. The publication in Figure 1 is a clear example of such a collaboration as the University of Maastricht is mentioned as an additional affiliation beside the address of Johnson and Johnson. While the respective address is clearly referring to a university, other addresses are more ambiguous requesting the need for a string-algorithm. Hence, I developed an algorithm that recognizes short strings within addresses and links it to particular sectors: university, hospital, industry, government and non-profit. Figure 2 provides an example list of addresses with their allocated sector. For this dissertation, I was interested in affiliations of academia referring to institutions with a non-profit, long-term view on research, mainly aiming to increase the world's knowledge base for the greater social good. As this characterization applies to the four sectors of university, hospital, government and non-profit, their respective affiliations are considered to represent academic institutions.



Neuronal differentiation is accompanied by NSP-C expression

By: Hens, J (Hens, J); Nuydens, R (Nuydens, R); Geerts, H (Geerts, H); Senden, NHM (Senden, NHM); Van de Ven, WJM (Van de Ven, WJM); Roebroek, AJM (Roebroek, AJM); van de Velde, HJK (van de Velde, HJK); Ramaekers, FCS (Ramaekers, FCS); Broers, JLV (Broers, JLV)

CELL AND TISSUE RESEARCH
Volume: 292 **Issue:** 2 **Pages:** 229-237
DOI: 10.1007/s004410051054
Published: MAY 1998
View Journal Information

Abstract

Neuroendocrine-specific protein (NSP) reticulons are expressed in neural and neuroendocrine tissues and cell cultures derived therefrom, while most other cell types lack NSP-reticulons. Three major subtypes have been identified so far, designated NSP-A, NSP-B, and NSP-C. We have investigated the correlation between the degree of neuronal differentiation, determined by morphological and biochemical criteria, and NSP-reticulon subtype expression. For this purpose, several human neuroblastoma cell lines, exhibiting different degrees of neuronal differentiation, were examined immuno(cyto) chemically. It became obvious that the expression of NSP-C, as detected by immunofluorescence microscopy and Western blotting, is most prominent in cell lines with a high degree of neuronal differentiation, such as LA-N-5. Such highly differentiated cells also express other neural and neuroendocrine markers, such as neural cell adhesion molecule (NCAM), neurofilament proteins, synaptophysin, and chromogranin. NSP-A was observed in all cell lines to a different extent. However, no clear correlation was observed with the degree of neuronal differentiation as defined by other neuronal and neuroendocrine markers or morphology. NSP-B could not be detected. The induction of neuronal differentiation with nerve growth factor, dbcAMP, and retinoic acid in the rat pheochromocytoma cell line PC12 and the human teratocarcinoma cell line hNT2, respectively, induced the expression of NSP-A and NSP-C in these cell lines parallel to the induction of neurofilament protein expression. It is concluded that NSP-C expression, in particular, is strongly correlated with neuronal differentiation.

Addresses:

+ [1] Univ Maastricht, Dept Mol Cell Biol & Genet, NL-6200 MD Maastricht, Netherlands
+ [2] Janssen Res Fdn, Dept Cellular Physiol Life Sci, B-2340 Beerse, Belgium

Figure 1. Example publication data, affiliation data to determine involvement of academia-industry.

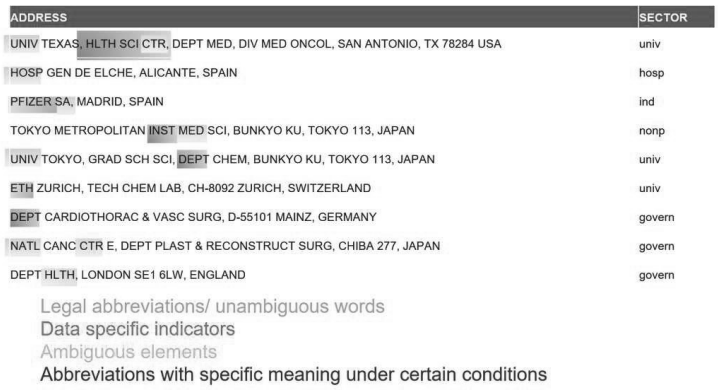


Figure 2. Example list of the applied string algorithm

Basic-applied research: CHI-journal classification

Within this dissertation, I will frequently refer to the basic and applied scientific research. Basic research has the aim to increase fundamental understanding of a phenomenon without a particular application in mind. In contrast, applied research has the clear intention to improve an existing application or find an application to solve a particular problem. Based on this distinction, basic research has been related to larger scientific freedom, broader search, higher risk and larger time investment (Stokes 1997). While this distinction and its consequences for the expectations and organization of science will be discussed in more detail within the first and second chapter of this dissertation, I will explain the empirical methodology used in this dissertation to make the basicness distinction: the CHI journal classification scheme.

The CHI journal classification scheme determines the basicness level of journals and has been developed by CHI Research Inc. The example presented in Figure 1 is a basic-research publication as the journal "Cell and Tissue Research" is categorized under basic research by this classification. As described in Hamilton (2003), the classification is based on the assumption that applied journals are more likely to cite basic journals than basic journals are to cite applied journals. For further refinement, the classification relies on expert classifications. In total, the classification identifies four levels of basicness: 1) clinical observation, 2) clinical mix, 3) clinical investigation, and 4) basic science. While the CHI classification has become the standard for identification of basic research thanks to its broad coverage and open access (e.g. Della Malva *et al.*, 2015; Belderbos, Gilsing and Suzuki, 2016; Arora, Belenzon and Sheer, 2017), it does have its limitations.

First, there is no detailed method description on the CHI classification available making it impossible to reproduce and update. As the last update runs back to 2006, this limitation has become of increasing importance. Fortunately, this is no concern for my studies because the data analyses only examine research activity until 2002. Some scholars have generated alternative classification methods, based on sector affiliations (Robert J W Tijssen 2010) or text mining (Lewison and Paraje 2004; Boyack et al. 2014).

Second, the CHI classification is a journal based classification scheme. This is an important limitation as articles within one journal, and even one volume, may differ significantly in their level of basicness. Hence, article-level classifications based on data-mining of abstracts and titles (Lewison and Paraje, 2004; Boyack *et al.*, 2014) may results in a more precise classification. They, however, require substantial investments of computation time and are highly dependent on the selected key words. In my studies, I aim to, at least, lower this concern by taking into account all versions of the CHI classification from 1973 to 2006 and reclassifying journals across time. Some journals, namely, shift their focus from basic to applied research or vice versa. One example is "Clinical And Experimental Hypertension" which was reclassified from level 2 tot level 3 in 1993. Nonetheless, a journal classification remains a rough approximation of the basicness level of an article especially within broad journals like "Science" (level 4).

A third disadvantage of the CHI-classification is its limited coverage. While the classification spans 7.506 journals across 16 Caspar research areas, small journals are often not classified. In our dataset, 4.3 percent of the journals were not included in the CHI classification. By assuming that most scientific journals and articles report on fundamental studies, I classified these journals under basic research. Finally, while the CHI classification spans a broad range of research areas, its internal validity within each area can be questioned. Social sciences and engineering sciences are for instance tougher to classify within four basicness levels. Moreover, the theoretical meaning of each level and of the concepts basic/applied research may differ considerable from area to area.

In the light of this dissertation, I performed two checks to test the correctness and investigate the meaning of the CHI classification within my research setting of the bio-pharmaceutical industry. First, I manually examined 26 articles of my publication sample of biotech and pharmaceutical companies. A comparison of the appointed classification with the journal's description and the article's abstract and title revealed that the articles of basicness levels 1 and 2 referred to characteristics of clinical trials phase 2 (e.g. efficacy, control group, placebo,

clinical practice). In contrast, articles of basicness levels 3 and 4 represented mainly discovery - lab tests and theoretical work - and some small-scale clinical trials of phase I. Second, I compared the CHI-classification to the publication classification of Pubmed (MeSH, medical subject headings) indicating clinical trials. Respectively 26 and 24 percent of the publications of basicness levels 1 and 2 have a MeSH term referring to clinical trials while this is only 5 and 0.4 percent for basicness levels 3 and 4. Based on the results of both exercises, it appears that within life sciences, basicness levels 1 and 2 roughly correspond to applied research with articles referring to clinical trials, while basicness levels 3 and 4 are significantly more likely (t-test: t= 186.8830, p= 0.0000) to represent basic research with in-vitro tests and fundamental discoveries.

Following the use of the CHI-journal classification, the basic-applied research distinction made within this dissertation follows the separation between laboratory or animal tests and clinical tests on humans. Furthermore, it implies that 'translation', which is discussed in Chapter 2, refers to the first area of translational research according to Woolf (2008): the process of applying discoveries generated during research in the laboratory, and in preclinical studies, to the development of trials and studies in humans. In other words, my research does not discuss the second-area of translational research: bench-to-bedside research focusing on improved access and patient-clinician relationship.

Within each of the following chapters, a more detailed description will be given of the particular dataset used for each individual study. Moreover, each study is positioned within its own relevant literature and its individual contribution is specified in the chapter's conclusion. The general conclusion provided at the end of this dissertation reflects on the dissertation's aim to provide insight in the scientific efforts that could form an effective treatment for particular causes of the productivity crisis. Moreover, the conclusion discusses some overarching limitations of the dissertation and provides some attractive directions for future research.

Chapter 1

Scientific Research, Drug Candidates and Drug Development

Who does the science and who develops?[1]

ABSTRACT

The pharmaceutical industry is not only one of the world's most research-intensive industries, its performance is also crucial for public health. However, the industry is suffering from a productivity 'crisis': new drug launches are limited despite soaring R&D expenditures. While firms are consistently patenting new drug candidates, most of these are not taken-up in (clinical) development projects, and only a few reach the market. Scholars and practitioners have referred to a lack of scientific understanding and the vertical disaggregation of the industry as possible root causes of the limited productivity of the pharmaceutical industry. In this study we examine the influence of scientific knowledge as prior art of patented drug candidates on the probability that firms decide to pursue clinical development. Examining the references to scientific literature of patented drug candidates of 131 biopharmaceutical firms (1995-2002), we confirm that the scientific background of these drug candidates is an important antecedent of take-up in clinical development. While the results indicate that overall both basic and applied scientific references increase the likelihood on development of patented drug candidates, basic research references performed by the patent owner appear to only stimulate the patent owner itself to take-on development. These findings raise questions on firms' rationality in dealing with the take-up decision and the complementarity of research and development activities with potential advantages of vertical integration.

[1] This chapter is based on a working paper co-authored by René Belderbos (KU Leuven, Maastricht University, UNU-Merit), Bart Leten (KU Leuven & U Hasselt) and Stijn Kelchtermans (KU Leuven & U Hasselt).

1.1 INTRODUCTION

During the last decades, the pharmaceutical industry is being confronted with numerous challenges resulting in the so-called 'productivity crisis': declining revenues, soaring R&D costs and, most importantly, limited numbers of new drugs that reach the market (I. Cockburn 2007; Kola and Landis 2004). In their quest to the root causes of this crisis, industry experts have called for a "return and renaissance to scientific depth and scrutiny in the understanding of human and animal physiology" (Kramer, 2016, p123). Moreover, scholars have been criticizing the vertical disaggregation of the industry leading to an organizational separation between research and development (Iain M Cockburn 2004; Arora and Gambardella 1994; I. Cockburn 2007).

As the 'productivity crisis' places firms under enormous pressure to effectively allocate funds, the decision on the development of drug candidates, i.e. whether to take them up into clinical trials (phase 1), is of crucial importance. Take-up entails substantial R&D costs (Dimasi, Hansen and Grabowski, 2003; Adams and Brantner, 2006; Arora *et al.*, 2009) under high uncertainty regarding development success and market potential. However, despite the importance of the clinical development decision, this decision process has not received much research attention.

In this study, we aim to provide some first insights into the drivers of the take-up decision, with a specific focus on the role of the scientific origins of patented drug candidates, as evidenced by scientific citations. Studies describing development failures have highlighted the importance of scientific insight in drug development (e.g. Kola & Landis 2004; Hay *et al.* 2014). First, the explorative and fundamental nature of basic scientific research may increase the expected market size, as the drug is more likely to cause a radical improvement in the treatment of a disease. Second, firms' expectations on whether the challenges of the clinical trial stage can be overcome will depend on the difficulties to translate insights from (basic) science to the clinical development process and the expected ability to solve impending challenges during development. In this regard, applied scientific insights related to the drug candidate may play an important role. As applied research is closely linked to the development process, it provides practical insights and prospects on difficulties during clinical trials. Furthermore, it may inform the firm about crucial techniques to solve impending applied scientific puzzles, such as devising an effective drug delivery mechanism. Hence, it is of interest to examine whether a patented drug candidate's scientific background in basic or applied research increases the probability of take-up in clinical development.

Related to the request for stronger scientific understanding (Kola and Landis, 2004; Barrenho, Miraldo and Smith, 2013; Cook et al., 2014; Hay et al., 2014; Waring et al., 2015; Chiou et al., 2016), questions have been raised concerning the organizational separation between (scientific) research and (drug) development in the industry. Some scholars have been emphasizing the importance of firm's engagement in both scientific discovery and development (Arora, Belenzon, and Patacconi 2015; I. Cockburn 2007). They state that the market for technology is imperfect due to the difficulty of transferring scientific understanding which is of a tacit and embedded nature (Iain M Cockburn 2004; Arora and Gambardella 1994; I. Cockburn 2007). Additionally, they have referred to the lemons problem indicating the difficulty of external parties to value inventions (Arora *et al.* 2009; Pisano 1997). In line with this argumentation, it can be expected that the firms who contributed to the scientific background of their drug candidate may be more inclined to develop internally as they have full access to the embedded and tacit knowledge and are not confronted with a lemons problem. More so, firms may be overoptimistic and overcommitted to drug candidates originating from internal scientific research resulting in the development of glorified candidates, by inferior developers (Natalicchio, Messeni Petruzzelli and Garavelli, 2014; Antons and Piller, 20*15; Peck et al.*, 2015; Eliëne *et al.*, 2018).

In this study, we aim to shed more light on these issues by examining the relationship between the scientific background of patented drug candidates and the likelihood that the drug candidate is selected for development in clinical trials, by the firm owning the patent, or by another firm. In the scientific background, we distinguish between basic and applied scientific research, and whether or not the firm itself contributed to the scientific research cited in the patent. We analyze the patents protecting potential drug candidates of 131 major firms in the biopharmaceutical industry worldwide between 1995 and 2002.

The results show an overall significant positive relationship between the scientific background of the drug candidate, either basic or applied, and the likelihood of take-up into the clinical development process. A firm is more likely to bring a drug candidate into the development process itself if it has contributed to the basic research background of the patent. Internal contributions to the applied research background of the patent appears unrelated to the decision between in-house or external development. Overall, we theorize that the patterns relating scientific background to drug development decisions are due to 1) the more fundamental insights on drug and disease properties when drug candidates draw on basic scientific research, which in turn may increase expected economic value (Teece 2010); 2) the articulated

information on translation and clinical development approaches that may be represented by a drug candidate's scientific background in applied research, improving the expected chance of clinical development success, 3) the greater complexity and tacitness of knowledge related to basic research, favoring intra-firm exploitation, 4) the high commitment and optimism regarding internal research projects, stimulating extended investments into development.

Our results confirm the importance of scientific insights, be it basic or applied, for drug discovery and development. In-house applied research appears to provide similar opportunities within the focal firm than on the 'market for technology', i.e. enabling the development of the drug candidate by other firms. In-house basic research, however, appears an important antecedent of drug development by the patent owner. Still, the underlying mechanism for this phenomenon remains unclear. On the one hand, firms who have performed basic scientific research regarding a drug candidate may be committed to extend their investments into clinical development even though the drug candidate may be of low potential or the firm may not be the most suited developer. On the other hand, the finding may suggest that there is complementarity between research and development hampering the effectiveness of the vertical division of labor in the global pharmaceutical industry. This would imply that it is primarily in-house basic scientific research that may give firms more opportunities in drug development and that firms reducing scientific research in-house may indeed be 'killing the golden goose' (Arora, Belenzon, and Patacconi 2015). While the supplementary analyses contradict the mechanism of irrational considerations, like over-commitment, they do not provide solid evidence of the mechanism of enhanced development capabilities. Hence, we encourage further research on the topic as it is of high economic and social importance that valuable scientific discoveries are translated into useful innovations.

1.2 IMPORTANCE OF SCIENCE FOR INNOVATION

The link between science and technology is a key topic within the innovation literature. Jaffe (1989) postulated a stimulating role of science in the patent production function of regions. This finding has been supported by studies at the level of the firm indicating that firms' scientific activity can be related to their innovation performance, mainly measured by citation weighted patent counts (Adams 1990; Gambardella 1992; Della Malva *et al.*, 2015; Fleming & Sorenson 2004). These papers indicate that scientific knowledge not only increases firms' ability to generate a higher quantity of patents but also to elevate the quality of their inventions. As

science provides a deeper understanding of the technological landscape in which firms search for inventions (Gambardella, 1992), it should help them prioritize research avenues and avoid costly research trials that lead to low-value outcomes (Fabrizio, 2009).

Since 1997, scientific non-patent references (SNPRs) are being analyzed to examine the link between science and technology in a more direct manner. The pioneering study of Narin *et al.* (1997) found an increasing reliance of U.S. patents on scientific research. More precisely, while the number of U.S. patents grew by 30 % from 1987 to 1993, the number of references to U.S. scientific publications tripled. Later, scholars have analyzed the association between SNPRs and the quality of their citing patent, measured by the patents' citation count. Not only are these studies limited in number, they also present diverging conclusions.

Fleming and Sorenson (2004) only found a significant positive effect of having a SNPR if the patent required coupling, meaning that it contained a highly interdependent combination of patent classes. In a study of patents across multiple industries, Cassiman and Veugelers (2006) found that the citations to scientific publications do not significantly explain the forward citations of the patent, though they are positively related to the scope of forward citations. An important note in this study is that they do find a weak impact of SNPRs on the patent citation count when limiting their sample to science intensive industries, like pharmaceuticals. In general, they state that the linkage to science at the firm level matters more for forward citations than the linkage at the invention/patent level. Sapsalis and colleagues (2006), who analyzed Belgian patents of all industries, even presented a negative significant impact of SNPRs on the number of patent citations especially when the scientific references are not published by the inventor. They suggest that this negative finding may be due to the public nature of these 'external' scientific publications entailing no competitive advantage and, hence, limited commercial value. The exploratory study of Sternitzke (2010) described the non-patent references of four different types of patents based on their technological novelty and market relevance: incremental innovation, technological breakthrough, market breakthrough and radical innovation. His findings indicate that references to publications of universities and research institutes are relevant for all types of innovation. Radical innovations, patents with both high technological novelty and market relevance, cite significantly more of these academic publications and cite more basic research. In contrast to his expectations, references to own publications (or patents) are more related to technological novelty than to market relevance. Finally, Zahringer and colleagues (2017) have indicated that inventions with highly-cited scientific references are more frequently cited by other patents. This limited set of studies

provides a much more nuanced view on the link between science and innovation, suggesting that additional research is required to uncover how science influences technology development.

In this study, we aim to contribute to the literature by positioning the science-technology link within the specific context of the pharmaceutical industry. We have chosen this specific industry for two reasons. First, the development process of inventions to marketable products is well documented within pharmaceuticals. This allows us to determine an alternative measure for technological quality: firms' decision to invest in the development of the invention, i.e. to take-up the drug candidate in clinical trials. We believe that it is important to investigate another determinant of technological quality as prior research had indicated that citations are only a noisy measure of patent value (Wagner and Wakeman 2015; Arts, Paolo Appio, and Van Looy 2013). Moreover, highly cited patents are not always related to successful development (Chi*ou et al.*, 2012; Wagner and Wakeman, 2015). In contrast, our measurement of 'take-up' forms the first step in the development process after the patent on a drug candidate is granted. This take-up decision will be based on the expected value of the drug candidate and will not be influenced by later stage events that are difficult to measure and control by firms, such as additional scientific discoveries, regulatory changes or market launches by rivals. Secondly, the pharmaceutical industry is experiencing a 'productivity crisis': an imbalance between scientific progress, R&D investments and new products launched on the market (DiMasi, Grabowski, and Hansen 2016; DiMasi, Hansen, and Grabowski 2003; Pammolli, Magazzini, and Riccaboni 2011). By examining the relationship between scientific references of patented drug candidates and take-up in clinical trials, we hope to shed more light on the role of science in the productivity crisis.

1.3 THE PRODUCTIVITY CRISIS IN THE PHARMACEUTICAL INDUSTRY

During the last decades, the number of new drug launches within the pharmaceutical industry has been dropping despite souring R&D expenditures (Cockburn 2007; Kola and Landis 2004). Scholars have been speculating on possible causes of this 'productivity crisis': 'low hanging fruits have been picked' (Danzon, Nicholson, and Pereira 2005; Crispeels, Willems, and Scheerlinck 2018; Moors, Cohen, and Schellekens 2014; Pammolli, Magazzini, and Riccaboni 2011), a lack of translational science (Kola and Landis 2004; Cockburn 2007) and the disaggregation of the value chain (Cockburn, 2007; Arora *et al.*, 2009). Nonetheless, the literature has been unable to identify the root causes of this 'crisis'.

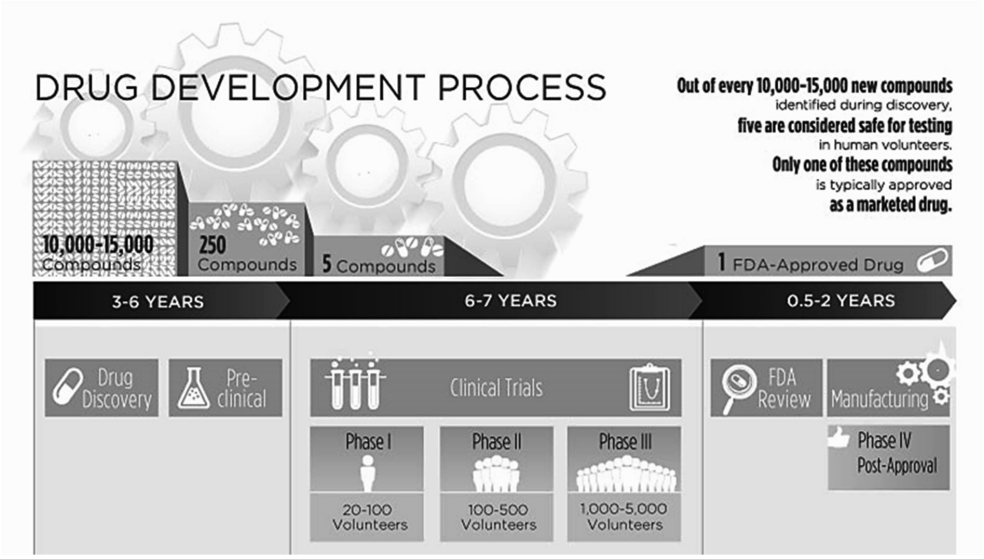


Figure 1.1. The drug development process
(source: http://www.ppdi.com/About/About-Drug-Discovery-and-Development)

Figure 1 represents the stage gate process of drug development. For every 250 compounds tested in laboratories and on animals, on average only five are taken up into clinical tests on humans (Campbell, 2005; PPD, 2017). Moreover, most of them never realize drug approval. Of the five that advance to clinical testing on healthy (phase I) and infected (phase II and phase III) humans, only one is eventually approved by the Food and Drug Administration (Campbell, 2005; PPD, 2017).

In this study we will focus on the first gate within the development process, where firms decide whether to take up the drug candidate in clinical trials (phase I). This is an important decision as the larger share of R&D costs is incurred at the clinical trial stage (Dimasi, Hansen and Grabowski, 2003; Adams and Brantner, 2006; Arora *et al*., 2009). The take-up decision will depend on firms' expectation concerning the costs of clinical development, the odds of clinical trial success, and the potential market size. The latter expectation increases when the drug entails a radical improvement in the treatment of a disease and is related to discoveries in its fundamental understanding. The firm's expectations on whether the challenges of the clinical trial stage can be overcome depend on the difficulties to translate insights from complex and tacit basic science to the clinical development process and the expected ability to solve impending applied scientific puzzles, such as devising an effective drug delivery mechanism.

In short, as Ceccagnoli and Jiang (2013), we believe that while the take-up decision is made early within the development process it is a key selection point in which firms take into account expected difficulties and benefits: development risks, knowledge transfer difficulties, complementary assets and expected returns on the drug market.

1.4 SCIENTIFIC BACKGROUND OF THE DRUG CANDIDATE

By describing development failures within the pharmaceutical industry in detail, multiple studies have highlighted a lack of scientific insight in the development of drug candidates leading to a mismatch between disease, compound, target group and clinical techniques (Kola and Landis, 2004; Barrenho, Miraldo and Smith, 2013; Cook *et al*., 2014; Hay *et al*., 2014; Waring *et al.*, 2015; Chiou *et al*., 2016). Additional to the well documented role of science in the creation of innovation (e.g. Adams 1990; Gambardella 1992; Della Malva *et al.* 2015), these studies signal that scientific insight may also play an important role in the development of innovations. For example, Paul and colleagues (2010) identified the doubtful validity of the target for the disease pathophysiology as a key reason for dropout of a drug candidate, besides inadequate compound characteristics or safety issues (Eichler *et al.*, 2009). Despite the aforementioned concerns over scientific productivity, these analyses of failures pinpoint exactly those issues where science can be expected to generate better prospects for drug development. Especially in the current age of 'industrialized' drug development, characterized by high throughput screening of chemical libraries, industry experts have called for a "return and renaissance to scientific depth and scrutiny in the understanding of human and animal physiology" (Schuhmacher *et al.* 2016, p 123).

1.4.1 Basic Scientific Research

Basic scientific research is known as a provider of fundamental insights, serving as a map of the technological landscape, guiding firms towards a global optimum (Fleming and Sorenson, 2004; Lim, 2004; Della Malva, *et al*., 2015). Reaching a global optimum entails a significant improvement in comparison to existing knowledge and, hence, the expectation on substantial returns after commercialization. Moreover, more fundamental discoveries may be more broadly applicable leading to multiple markets in which value can be generated. For drug candidates with a basic-research background, this has two implications. First, the broad applicability implies that firms are more likely to identify a development trajectory within their strategic aims and resources. Second, the larger generalizability may lead to drug candidates for an extend product market whit high expected returns after successful commercialization.

However, the impact of basic research on the expected development success of a drug candidate is less clear. Drug candidates pushed by theoretical advancements may entail the exploration of new directions leading to more risk in how the lab findings will translate to the human body. In contrast, field experts have indicated that finding the 'right' fit between target and disease and setting up the 'right' clinical trials depends upon fundamental understanding of the human body, animal models and disease essays (Kola and Landis 2004; Waring *et al.* 2015; Barrenho, Miraldo, and Smith 2013; Cook *et al.* 2014; Chiou et al. 2016; Hay *et al.* 2014). Also, innovation scholars have referred to basic research as a guide to the most fruitful directions in research, avoiding costly clinical trials (Fabrizio 2009; Fleming and Sorenson 2004; Colen et al. 2018). In short, basic research may increase the expectation of a large market and provide important insights on the right approach to development, though, it may also increase the development challenges in capturing this market.

1.4.2 Applied Scientific Research

In contrast to the view that theoretical change (basic research) underlies technological innovation (Adams 1990), innovation may also build upon prior development experiences or, in other words, applied research (Rosenberg 1976, Schmitz 1989)2. In the context of bio-pharmaceuticals, this includes reports on clinical trials and translational research from which firms may extract insights on crucial techniques and methods to solve impending applied scientific puzzles. A case in point is the progressive insight on the importance of transport mechanisms used by a drug to penetrate through biological membranes. While this was long considered to be a merely biophysical problem that could be dealt with later in the development process, there is mounting evidence that knowledge about transport proteins in a candidate drug strongly improves its odds of success (e.g. Dobson & Kell 2008). A study by Pfizer of 44 internal development programs reaching clinical trial phase II between 2005 and 2009 confirmed the importance of knowledge of the pharmacokinetic behavior for ultimate success in drug development (Morgan et al., 2012). Moreover, as applied research is closely linked to the development process, it provides practical insights and prospects on difficulties during clinical trials. Most importantly, reports on clinical trial outcomes contain important information on the complex interactions within the human body, which are highly difficult to

[2] As indicted in the introduction, we refer to applied research as publications reporting on the development of trials and studies in humans. Hence, this study does not discuss bench-to-bedside research focusing on improved access and patient-clinician relationship.

foresee from a theoretical perspective or pinpoint within a lab environment. Consequently, the presence of applied scientific understanding of a drug candidate may enlarge the expectations on success in clinical development. However, in contrast to basic research, applied research is related to the exploitation of existing markets and learning from experience which are path-dependent processes leading to local optimums with limited expected return after successful development (Fleming and Sorenson 2004).

1.5 THE MARKET FOR TECHNOLOGY IN THE PHARMACEUTICAL INDUSTRY

Since the start of biotechnology, the structure of the pharmaceutical sector has been changing (Cockburn, 2004, 2007; Stuart, Ozdemir and Ding, 2007; Arora *et al*., 2009; Malik, 2011; Belderbos, Gilsing and Suzuki, 2016). While traditionally, pharmaceutical companies were vertically integrated firms active in both research and development, the value chain has disaggregated with some players focusing on research, others on development and a few remaining highly integrated. Complementary, an active market for technology has been established with increasing vertical transactions like licensing of drugs and drug candidates, joint ventures and strategic technology alliances, and mergers and acquisitions (Malik 2011). These transactions imply that patented drug candidates are not necessarily further developed in clinical trials by the (original) patent owner, but are often sponsored by other firms. Moreover, firms view on the market for technology has changed considerably. While there was traditionally a reluctance to 'give away' internal inventions and the market of technology was used to transfer internally unused technology, currently licensing and related practices are not merely an afterthought but an active part of firms' innovation strategy (Lichtenthaler 2011).

Past research has indicated that financial revenues (fixed fees or royalty payments) are most often the reason for firms to out-license their innovations (Arora and Gambardella 2010; Gambardella, Giuri, and Luzzi 2007; Sakakibara 2010; Zuniga and Guellec 2008; Radauer and Dudenbostel 2013). Additionally, out-licensing can increase the demand for a firm's technology (Shepard 1987), can be a mechanism for capturing a share of future inventions of the licensee (Ahuja, Lampert, and Novelli 2013), or can enable the company to enter new markets (Lichtenthaler 2011). Underlying these monetary and non-monetary motives is the lack of necessary complementary assets at the patent owner to develop or market the patented invention and acquire the highest returns (Lichtenthaler 2011).

In line with this reasoning, the division-of-labor theory states that the market for technology should stimulate innovation performance. Scientific research and product development are two distinct tasks with their own needs for specific resources and capabilities. Science, being risky and unpredictable, requires an open and flat company structure with strong interactions with the academic community. Development, being more standardized, is subject to scale and scope economies and best pursued in large, hierarchically organized firms. Some evidence for these assertions is given by the finding that firms appear unable to produce both high quality publications and highly cited patents (Gittelman and Kogut 2003). Following this theory, patents would be taken up by the most capable developer. Chesbrough and Chen (2013) emphasize the importance of licensing out drug candidates that can turn out to be successful within other firms by describing the example of Cubicin. This well-known antibiotic reaching sales over 800 million USD in 2012, is a good illustration of how licensing activity can be crucial for development success. Moreover, by studying patent licensing agreements Laursen and colleauges (2010) and Leone & Reichstein (2012) showed that licensees are faster at innovating than non-licensees.

However, in light of the productivity 'crisis', scholars have been questioning the effectiveness of this disaggregation of the value chain within pharmaceuticals (Cockburn 2004; Arora and Gambardella 1994; Cockburn 2007). Division of labor is optimal in perfect markets with complete information, minimal transaction costs and fluid transfer of assets. In contrast, pharmaceutical R&D is characterized by long-term, risky projects entailing complex and sticky knowledge, difficult to monitor activities, substantial project-specific unrecoverable investments and vertically complementary outcomes (I. Cockburn 2007). Because of these characteristics, licensing of drug candidates is subject to numerous difficulties: difficult knowledge transfer, adverse selection bias, entanglement with core IP, product liability risk and exposure risk of key assets (Chesbrough and Chen Eric L. 2013; Ceccagnoli and Jiang 2013). Also in the studies mentioned above important conditions for the higher innovation performance of licensees have been identified. Leone and Reichstein (2012) showed that licensees are faster at innovating than non-licensees, provided that contracts do not contain grant-back clauses (which allow licensors to license back new innovations that build further on licensed technologies) and licensees are familiar with licensed technologies. Laursen *et al.* (2010) indicated that gaining access to unfamiliar, distant technologies in technology markets requires prior investments at the firm level in monitoring abilities. These conditional findings suggest difficulties in development when background knowledge on the innovation is lacking.

Some studies have then also emphasized the need for close interaction between the licensee and licensor to enable knowledge transfer (Thursby and Thursby 2011; Agrawal 2006; Klueter, Monteiro, and Dunlap 2017). Similarly, Ceccagnoli and Jiang (2013) have noted the importance of knowledge transfer capabilities at the licensor and absorptive capacity at the licensee, especially in industries where research and development are complementary. In conclusion, this theoretical view and the first empirical indications speculate that vertical integration strategies of firms engaged in scientific research as well as clinical development may bring productivity advantages.

To disentangle the above-mentioned effects, we will distinguish between development by the patent owner and by other firms when analyzing the effect of different scientific backgrounds in patented drug candidates. While above, we already made the distinction between basic and applied research. The former entails studies and insights of a fundamental and more general nature that tackle complex and risky problems. It typically requests deep learning to comprehend the issue and long-term investments to obtain results. When the basic research background of a drug candidate originates from internal scientific activities of the firm, this may influence firms' decision to further develop internally.

We expect that drug candidates with internal scientific research as prior art are more likely to lead to take-up by the patent owner for reasons. First, the more general nature of basic research may lead to drug candidates with potential in multiple development projects causing greater concerns about intellectual property. Furthermore, the tacit nature of basic research reinforces problems of knowledge transfer to other firms, such that there are advantages in drawing on the tacit and complex scientific background in internal development. Although partner-embedded licensing deals may be set up to deal with knowledge transfer issues (Klueter, Monteiro, and Dunlap 2017; Thursby and Thursby 2011), these close interactions also have drawbacks: they may generate unintended knowledge spill-overs as partner firms access key scientists and development processes of the firm. Hence, drug candidates with an internal basic research background may be regarded less suited for licensing and clinical development by other firms. Third, the generalizability of basic research may increase the probability that at least one development trajectory of the drug candidate falls within the strategic aims and technological capabilities of the focal firm. Fourth, as basic research is lengthy and profound, the decision process may be influenced by irrational considerations. Its researchers and managers may be highly sensitive to the not-sold-here syndrome (Antons and Piller 2015; Natalicchio, Messeni Petruzzelli, and Garavelli 2014), optimism bias (*Peck et al.*, 2015) and over-commitment

(Eliëns et al. 2018). Moreover, both the decision of no take-up as take-up by other firms may be seen as a failure of the research team, which was actively investigating the fundamentals of the drug. Managers may want to avoid this negative annotation to the team and may give them the chance to proof their capabilities (Peck et al. 2015). Finally, the take-up decision may be biased by sunk-cost fallacy and a lack of proper consideration of opportunity costs (Peck et al. 2015). Consequently, drug candidates with a basic internal research background may be more likely to receive extended investments into development by their patent owner.

These arguments are less likely to hold if the patented drug candidate cites the firm's applied scientific research. Not only is applied research directed to a specific application, treatment or patient group limiting IP issues. It is also of a more exploitative nature building on existing knowledge and processes requesting less time and project specific investments. Hence, the take-up decision for drug candidates with an internal applied research background may be less influenced by irrational considerations. Finally, as applied research is less theoretical and more codified reporting on past clinical trials and particular methods, the presence of some underlying applied research may instead reassure licensees of the drug candidate's potential and the possible directions for clinical development, minimizing the risk of 'buying lemons' (adverse selection bias) (Pisano, 1997b; Arora *et al.*, 2009; Natalicchio, Messeni Petruzzelli and Garavelli, 2014). Similarly, Ceccagnoli and Jiang (2013) have referred to supplier's knowledge transfer capability as an important enabler of licensing. They stated that codified information, like experiences with failed experiments or know-how about the environments in which the technology works best, will increase the likelihood that potential buyers will understand the value of the suppliers' technology and will facilitate the integration of the technology within the particular needs and setting of the buying company.

1.6 DATA AND DESCRIPTIVE STATISTICS

This study focuses on 131 of the largest R&D spending (in absolute terms) pharmaceutical and biotech companies with headquarters in the United States, the EU or Japan.[3] As company structures change regularly within the pharmaceutical industry, we analyzed the firms' annual reports for listed subsidiaries resulting in a rigorous consolidation for eight consecutive years,

[3] As reported in the 2004 EU Industrial R&D Investment Scoreboard. This ranking lists the top 500 corporate investors in R&D whose parent is located in the EU, and the top 500 companies whose parent is located outside the EU (mainly US and Japan), based on corporate R&D expenditures in 2003.

1995-2002. For the consolidated firms, we gathered complete patent and publication data. Additionally, we have information on the drug development projects, which we have linked to the firms and their patents. In the following, these three data sources are explained in more detail.

1.6.1 Patent Data

Patent data is extracted from PATSTAT (2013) at docdb family level (Bakker *et al*, 2016). During patent retrieval, we consider first-time patent applications by the consolidated company from 1995 to 2002. This first application year will function as reference year (year t) for calculating the control variables. The 18,363 patents under investigation are identified as granted primary patents, protecting a compound. We focus our attention to primary patents for two reasons. First, not all patents are established with commercial intent. Strategic grounds can underlie patent application: patent fencing, surrounding, blanketing or flooding (Thumm 2004; Blind, Cremers, and Mueller 2009). Second, while a development process originates from a primary patent, along the road to market approval, additional patents are being filed on techniques, treatments and materials. To mitigate both these issues, we restrict our patent sample to compound patents which are most likely applied for with commercial intent and invented within discovery or preclinical research at the start of the development process. We attempt to capture compound patents by restricting our sample to IPC classes A61, indicating medical use, and, more importantly, C07 or C12, representing respectively the organic chemistry and biotech compounds. The WHO confirms that most pharmaceutical products can be found under A61K, A61P and/or C07 (World Health Organization 2010). Additionally, a report to the WIPO additionally indicated the importance of C12 classes for identification of biotechnology and the broader A61 class to also include medical technology beside pharmaceutical applications (A61K) or therapies (A61P) (Schmoch 2008). Within the patents of our sample firms, the chosen IPC3 classes (A61 and C07) appear strongly related to appearance within the project data representing 94 and 88 % of the project patents. Patents in class C12 are less common within the project data (21%), indicating the still inferior position of biotechnology within pharmaceuticals in the period 1995-2002. Among the 48,112 granted patents of our sample firms 38 percent (18,363 patents) have an IPC class in A61 and in C07 or C12 forming our sample of drug candidates.

1.6.2 Project data

We link the observed patents to project data, the Pharmaceutical Industry Database (PHID) maintained at the IMT Institute for Advanced Studies in Lucca (Italy). While the patents under

investigation are applied for in the period 1995-2002, the project data allows use to track project activity up until 2013. So even with the long development time within pharmaceuticals, we are confident to observe the majority of development activity for each of our drug candidates. This database on pharmaceutical R&D projects is based on information from governmental agencies, industry conferences, press releases and contact with firms[4]. We regard a patent as taken-up into a development project when the patent is mentioned at least once within the project data as protecting a compound included in clinical research. The project database also reports on the lead company performing the clinical trials. We distinguish between internal and external take-up into a development project, when the lead company is respectively the patent owner or another company. To make this distinction, the complete company structure of the focal firm, including mergers and acquisitions, is considered during the full length of the patent's development project. Consequently, when inventor companies are acquired leading to the development of their patents by the mother firm, these developments are regarded as internal.

1.6.3 Publication data

By analysing SNPRs, we aim to have a more detailed view on the link between science and the commercial opportunities of innovation. Surveys of and interviews with patent inventors have shown that only a limited set of the cited references are unknown to inventors (Fleming and Sorenson 2004; Robert J.W Tijssen 2001; Callaert, Pellens, and Van Looy 2013). Although studies on the meaning and use of SNPRs discourage the interpretation of SNPRs as a direct knowledge source of the patented invention[5], they do signal that they can be interpreted as the relevant scientific background of the patent (M. Meyer and Persson 1998; Callaert, Pellens, and Van Looy 2013). More precise, while SNPRs do not necessarily form the origin of the inventive idea, they represent sources of relevant knowledge to explain and understand the invention, which will be instrumental in estimating the success of clinical trials and the potential commercial value of the drug candidate. Within our sample of patents, we link the non-patent

[4] This data is highly comprehensive. It includes full text entries comprising more than 200,000 patent applications since the early 1970s from the US Patent and Trademark Office, the European Patent Office and the World Intellectual Property Organization; detailed information about R&D projects spanning more than 28,000 compounds; 20,000 collaborative agreements; and sales figures on ~160,000 pharmaceutical products (both branded and generics) sold in the major markets between 1996 and 2008 (Pammolli et al. 2011). In other words, the project data has a highly comprehensive coverage.

[5] With the notable exception of Tijssen and colleagues (2000) who do indicate a genuine, direct empirical link between scientific NPRs and patented inventions.

Table 1.1. Descriptive statistics on SNPRs within the total set and subgroups of taken-up of drug candidates

	Total		**Not taken-up**		**Taken-up**		**Taken-up By other firms**		**Taken-up By patent owner**	
N	18363		17822		541		254		287	
	Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
D(SNPRs)	0.64	0.48	**0.64**	**0.48**	**0.78**	**0.41**	0.76	0.43	0.80	0.40
No. SNPRs	7.59	15.36	**7.44**	**15.06**	**12.45**	**22.70**	13.41	26.02	11.60	19.30
D(basic SNPRs)	0.63	0.48	**0.63**	**0.48**	**0.77**	**0.42**	0.74	0.44	0.79	0.40
D(basic internal SNPRs)	0.22	0.42	**0.22**	**0.41**	**0.33**	**0.47**	**0.27**	**0.45**	**0.38**	**0.49**
No. basic internal SNPRs	0.59	1.77	**0.57**	**1.73**	**1.11**	**2.64**	1.00	2.70	1.20	2.59
D(basic external SNPRs)	0.60	0.49	**0.60**	**0.49**	**0.73**	**0.44**	0.72	0.45	0.75	0.43
No. basic external SNPRs	6.33	13.22	**6.22**	**13.00**	**9.97**	**18.66**	10.80	21.28	9.23	15.99
D(applied SNPRs)	0.21	0.41	**0.21**	**0.41**	**0.35**	**0.48**	0.35	0.48	0.36	0.48
D(applied internal SNPRs)	0.02	0.15	**0.02**	**0.14**	**0.05**	**0.21**	0.05	0.21	0.05	0.22
No. applied internal SNPRs)	0.03	0.29	**0.03**	**0.28**	**0.08**	**0.38**	0.08	0.43	0.07	0.34
D(applied external SNPRs)	0.21	0.40	**0.20**	**0.40**	**0.34**	**0.47**	0.34	0.47	0.34	0.47
No. applied external SNPRs	0.63	2.14	**0.61**	**2.08**	**1.30**	**3.65**	1.52	4.27	1.10	2.99

T-test comparisons of variable means are performed for Not taken-up and Taken-up, and for Taken-up by other firms and Taken-up by patent owner.
Values indicated in bold are significantly different at the 0.05 level.

references to the Web of Science through a text-mining algorithm developed at the MSI department of the KU Leuven, resulting in a list of SNPRs. We recognize own SNPRs as those published by the patent owner[6]. Furthermore, we distinguish between basic and applied research by means of the CHI journal classification: levels 1 and 2 indicating applied research and levels 3 and 4 indicating basic research (Hamilton, 2003; Thursby and Thursby, 2011; Colen *et al.*, 2018). As we recognize the imprecision of SNPR data to quantify the importance of one reference against another, we apply dummy variables instead of counts to indicate the presence of a scientific background within the set of SNPRs.

1.6.4 Descriptive statistics

Table 1 shows the difference in SNPRs within different groups of drug candidates: the total set, those not taken up, those taken up, those taken up by the patent owner, those taken up by other firms. Of the 18,363 patents under investigation, 287 are taken up by the patent owner and 254 by other firms. This take-up rate (3%) is similar to estimations from previous studies (2%, as reported by Campbell 2005; PPD 2017). 64 percent of the sample patents include SNPRs, and cite on average seven scientific publications. Almost double this amount is observed for patents that are taken up into development. Patents are generally more likely to cite basic (63%) than applied research (21%). Most patents that cite applied research also refer to basic research: 95 % within all SNPRS, 72% within references to internal research. For both basic and applied research, patents refer more often to externally sourced knowledge. Finally, patents taken up by the patent owner are more likely to cite internal basic research though their tendency to cite other research is similar to patents taken up by other firms. As discussed in the data description, our analyses will use variables indicating the presence of basic and applied SNPRs (Table 4 Model 2) and basic internal SNPRs, applied internal SNPRs, basic external SNPRs and applied external SNPRs (Table 4 Model 3 and Table 5 Model 4a & 4b).

In Table 2, the control variables are listed with their mean and standard deviation across the different groups of drug candidates. The patent specific controls consist of the number of firms

[6] If the patent is co-owned by multiple firms, the SNPRs are considered as internal when at least one of the owners published the work. Only 62 of the 18,363 patents under investigation are owned by multiple sample firms in its first application year.

Table 1.2. Descriptive statistics on the control variables within the total set and subgroups of taken-up drug candidates

	Total		**Not taken-up**		**Taken-up**		**Taken-up By other firm**		**Taken-up By patent owner**	
N	18,363		17,822		541		254		287	
	Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
PATENT LEVEL CONTROLS										
No. of owners	1.12	0.34	1.12	0.34	1.12	0.34	1.12	0.34	1.11	0.35
Scope: # IPC6 classes	8.97	6.16	**8.84**	**6.05**	**13.42**	**8.04**	12.97	8.67	13.81	7.44
No. of inventors	4.46	3.17	**4.43**	**3.14**	**5.44**	**3.67**	5.52	3.26	5.36	4.00
FIRM LEVEL CONTROLS										
Log(R&D $_{\text{in million \$, in t}}$)	5.94	2.20	**5.95**	**2.19**	**5.60**	**2.40**	**5.36**	**2.41**	**5.81**	**2.37**
Science intensity $_{\text{from t-4 to t-1}}$	1.40	1.41	1.40	1.40	1.50	1.65	1.48	1.82	1.51	1.48
Basic research focus $_{\text{from t-4 to t-1}}$	0.77	0.13	**0.77**	**0.13**	**0.78**	**0.15**	0.77	0.17	0.79	0.12
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.12	0.58	**0.11**	**0.57**	**0.19**	**0.74**	0.19	1.01	0.19	0.37
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$ ^2	0.36	12.09	0.35	12.18	0.59	8.76	1.05	12.72	0.18	1.20
Size of the market for techn.	209.18	51.66	**209.35**	**51.82**	**203.63**	**45.60**	201.71	46.59	205.33	44.72
Share within the market for techn.	0.01	0.02	**0.01**	**0.02**	**0.02**	**0.02**	**0.01**	**0.02**	**0.02**	**0.03**

T-test comparisons of variable means are performed for Not taken-up and Taken-up, and for Taken-up by other firms and Taken-up by patent owner. Values indicated in bold are significantly different at the 0.05 level.

owning the patent7, the number of inventors to control for team size and the technological scope of the patent (number of IPC6 classes). Additional to the figures presented in this table, the data shows that patents that cite basic research are more general. They have significantly more IPC6 classes, both in comparison to all other patents (t-test: diff -1.9, t -19.85, p 0.0000) and to those with at least one SNPR (t-test: diff -2.0, t -4.54, p:0.0000). Moreover, when the basic research is internal we see that the patents are of additional generality in comparison to patents only citing external basic research (t-test: diff -2.28, t-test -7.26, p 0.0000). These descriptive statistics support our theoretical argumentation that basic research is of a more fundamental nature leading to drug candidates that are more general and have the potential to provide financial returns in multiple commercial markets. Patents that are taken-up appear to have a larger inventor team and cover a broader set of IPC classes.

Table 2 also presents the control variables for the patent owner measured at the year of application of the patent. If multiple firms co-own the patent, these controls are taken at the mean. The size of the firm owning the patent is approximated by the natural logarithm of the firm's R&D expenditures. Drug candidates that are taken up appear to originate from significantly smaller firms, although larger firms develop more in-house. Second, we include the firm's scientific research intensity, measured by the total number of publications in the past four years divided by R&D expenditures in application year. Additionally, we control for the basic research focus of the firm, calculated by the share of basic research in its publication portfolio in the four years prior to application. Fourth, the linear and quadratic term of the firm's drug pipeline is taken into account, scaled by R&D expenditures. We may expect a reverse U-shaped relationship between a firm's pipeline and the tendency to develop their own drug candidates. When a firm has more recent clinical trial experience in ongoing projects, it may be more likely to have the capabilities in clinical trials. Moreover, a small pipeline may put profits in jeopardy pushing firms to take up more projects, possibly of lower quality and higher risk (Kola and Landis, 2004; Arora *et al*., 2009). In contrast, when a firm has a well-filled pipeline, it may be more selective in taking up new development projects. We also include two controls variables for the market for technology: the number of rival firms with projects in the relevant ATC classes (the number of potential candidates to license or sell a patent to) and the firm's

[7] We can only identify patent owners when they are included in a sample of 131 biotech and pharmaceutical companies (2004 EU Industrial R&D Investment Scoreboard).

Table 1.3. Correlation matrix of the regression variables

Categorical (take-up)	1.000														
Dummy (take-up)	**0.943**	1.000													
No. of owners	-0.004	-0.004	1.000												
Scope	**0.123**	**0.123**	0.008	1.000											
No. Inventors	**0.050**	**0.053**	**0.083**	**0.194**	1.000										
Log(R&D)	-0.020	**-0.026**	**0.091**	**0.045**	**0.048**	1.000									
Science intensity	0.012	0.010	**-0.022**	-0.020	**-0.034**	**-0.082**	1.000								
Basic focus	0.018	0.012	0.011	-0.013	0.016	**-0.121**	**0.271**	1.000							
Pipeline/R&D	**0.021**	**0.022**	-0.006	0.010	-0.005	**-0.089**	**0.100**	-0.004	1.000						
Size of market for technology	-0.016	-0.017	**-0.056**	**0.134**	**0.041**	**0.098**	**-0.082**	**-0.052**	-0.008	1.000					
Share in market for technology	**0.046**	**0.027**	**-0.037**	**0.067**	**-0.065**	**0.324**	**-0.107**	**-0.144**	**0.062**	**-0.039**	1.000				
D(basic SNPRs)	**0.051**	**0.049**	-0.011	**0.142**	-0.020	**-0.093**	**0.076**	**0.024**	0.019	0.002	**-0.071**	1.000			
D(applied SNPRs)	**0.057**	**0.061**	**-0.022**	**0.095**	-0.012	**-0.062**	**0.063**	-0.002	0.017	0.015	**-0.036**	**0.340**	1.000		
D(basic internal SNPRs)	**0.050**	**0.043**	**0.052**	**0.055**	-0.005	0.001	**0.166**	**0.086**	-0.002	-0.008	0.010	**0.409**	**0.277**	1.000	
D(applied internal SNPRs)	**0.030**	**0.030**	0.014	0.006	-0.014	-0.005	**0.060**	0.004	-0.002	0.001	0.001	**0.093**	**0.287**	**0.179**	1.000
D(basic external SNPRs)	**0.046**	**0.046**	**-0.024**	**0.153**	-0.015	**-0.108**	**0.075**	**0.024**	**0.022**	0.003	**-0.083**	**0.939**	**0.350**	**0.290**	**0.083**	1.000
D(applied external SNPRs)	**0.054**	**0.057**	**-0.025**	**0.100**	-0.012	**-0.061**	**0.057**	-0.004	0.019	0.014	**-0.036**	**0.336**	**0.980**	**0.262**	**0.180**	**0.351**

The correlations indicated in bold are significant at the 0.05 level. The shaded cells represent variables that are never paired within the same regression model.

share in the industry's ongoing development projects in the relevant ATC classes (the firm's relative experience and capabilities). We determine the relevant ATC classes for a patent by calculating the Jaccard Similarity between ATC-classes and IPC-classes across patents in the full project database. These variables form important controls in our analysis as they reflect the relative development capabilities of the focal firm and, hence, approximate licensing considerations of complementary assets (manufacturing facilities, marketing) and relevant clinical trial experience.

Finally, the models include dummies for time, therapeutic areas represented by the subclasses of IPC class A61P, and (4) chemical compounds (C07) and (5) biotechnical compounds (C12). By including these dummy variables the model takes into account the heterogeneity in development success across therapeutic areas (e.g. Mullard 2016; Kola & Landis 2004), the quality level of the existing drugs on the market, and the cost of clinical trials in the disease area. The correlation matrix of all the variables shown in Table 3 indicates no issues of multi-collinearity.

As our dependent variable is binary (whether a patent is taken up in clinical trials or not) we estimate logit models. In the analyses in which we also distinguish whether the patent owning firm or another firm takes up the patent in clinical development, there are three categories and hence we estimate multinomial logit models. The results tables will show both the coefficients, where the use of different reference groups shows whether there are significantly different effects between groups, and the average marginal effects (A.M.E.).

1.7 EMPIRICAL RESULTS

Table 4 presents the results of three logit regressions relating the probability that a patented drug candidate is taken up in a development project to the types of SNRPs of the patent.[8] Model 1 contains only the control variables. Patents with a broader scope and a larger inventor team are related to a higher probability of take-up. Broader patents may result in a divers set of application opportunities, increasing the likelihood that a commercially profitable opportunity can be found within the industry. Larger teams may generate more valuable patents as they combine a more divers set of capabilities, knowledge and points of view. Additionally, drug candidates of smaller (R&D budget) and more active (size of pipeline) and more capable (share

[8] Table A1 in appendix presents the average marginal effects of these regressions.

Table 1.4. Logit regression on the probability of take-up into clinical trials: Coefficients

	Model 1	**Model 2**	**Model 3**
	Controls	**Basic-Applied**	**Internal-External**
	Coef. (S.E.)	Coef. (S. E.)	Coef. (S.E.)
D(basic SNPRs)		0.58***	
		(0.12)	
D(applied SNPRs)		0.64***	
		(0.10)	
D(basic internal SNPRs)			0.31***
			(0.11)
D(applied internal SNPRs)			0.55**
			(0.24)
D(basic external SNPRs)			0.46***
			(0.12)
D(applied external SNPRs)			0.50***
			(0.11)
No. of owners	-0.16	-0.13	-0.16
	(0.15)	(0.15)	(0.15)
Scope: # IPC6 classes	0.08***	0.07***	0.07***
	(0.01)	(0.01)	(0.01)
No. of inventors	0.05***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)
Log(R&D $_{\text{in million \$, in t}}$)	-0.10***	-0.09***	-0.09***
	(0.02)	(0.02)	(0.02)
Science intensity $_{\text{from t-4 to t-1}}$	0.04	0.02	0.00
	(0.03)	(0.03)	(0.03)
Basic research focus $_{\text{from t-4 to t-1}}$	0.38	0.51	0.45
	(0.42)	(0.42)	(0.43)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.34**	0.29*	0.30*
	(0.16)	(0.17)	(0.17)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)
Size of market for technology	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Share in market for technology	7.60***	8.31***	7.93***
	(2.20)	(2.25)	(2.27)
Year dummies	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES
Biotech compounds dummy	YES	YES	YES
Chemical compounds dummy	YES	YES	YES
Constant	-3.32***	-3.82***	-3.63***
	(0.49)	(0.51)	(0.51)
LL	-2218.49	-2172.95	-2171.35

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.5. Multinomial Logit regression results explaining drug candidates not being taken up (1), taken up by other firms (2) and taken up by patent owner (3): Coefficients

	Model 4a		**Model 4b**	
Reference category	**1.Not taken up**		**2.Taken up by other firms**	
	2.Taken up by other firms	3. Taken up by owner	1. Not taken up	3. Taken up by owner
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.04	0.54***	-0.04	0.50**
	(0.16)	(0.14)	(0.16)	(0.21)
D(applied internal SNPRs)	0.69**	0.50	-0.69**	-0.19
	(0.32)	(0.32)	(0.32)	(0.44)
D(basic external SNPRs)	0.37**	0.54***	-0.37**	0.18
	(0.16)	(0.16)	(0.16)	(0.23)
D(applied external SNPRs)	0.58***	0.39***	-0.58***	-0.19
	(0.15)	(0.15)	(0.15)	(0.21)
No. of owners	-0.11	-0.11	0.11	-0.00
	(0.20)	(0.21)	(0.20)	(0.29)
Scope: # IPC6 classes	0.08***	0.07***	-0.08***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.06***	0.06***	-0.06***	-0.00
	(0.01)	(0.02)	(0.01)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.06*	-0.08**	0.06*	-0.02
	(0.03)	(0.04)	(0.03)	(0.05)
Science intensity $_{\text{from t-4 to t-1}}$	0.01	-0.04	-0.01	-0.05
	(0.06)	(0.05)	(0.06)	(0.07)
Basic research focus $_{\text{from t-4 to t-1}}$	-0.44	2.43***	0.44	2.87***
	(0.43)	(0.87)	(0.43)	(0.96)
Size of Pipeline $_t$/R&D$_{\text{in million \$, t}}$	0.19	1.58***	-0.19	1.39***
	(0.15)	(0.38)	(0.15)	(0.41)
Size of Pipeline $_t$ /R&D$_{\text{in million \$, t}}$^2	-0.01	-0.39**	0.01	-0.38**
	(0.01)	(0.16)	(0.01)	(0.16)
Size of market for technology	-0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	-19.24***	18.34***	19.24***	37.58***
	(7.17)	(2.66)	(7.17)	(7.62)
Year dummies	YES	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES	YES
Biotech compounds dummy	YES	YES	YES	YES
Chemical compounds dummy	YES	YES	YES	YES
Constant	-3.43***	-6.54***	3.43***	-3.11***
	(0.55)	(1.02)	(0.55)	(1.15)
LL	-2477.48		-2477.48	

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

in market for technology developers) are more often taken up in development. Finally, the pipeline shows only a linear relationship to take-up indicating no saturation effect.

The second model in Table 4 shows that drug candidates that cite scientific research are more likely to be taken-up into a development projects. Moreover, the coefficients of the two dummy variable, representing basic and applied research, are not significantly different (Chi2= 0.11, Prob.>Chi2 = 0.7402). The third model distinguishes internal and external scientific research, resulting in four dummy variables of interest. All types of SNPRs have a positive and significant effect on take-up, with no significant differences between the coefficients (Chi2= 1.87, Prob.>Chi2 = 0.5997).

Table 5 presents the multinomial logit results of analyzing the take-up decision: (1) not taken up, (2) taken up by other firms and (3) taken up by patent owner[9]. As the coefficients represent the increased likelihood of one categorical outcome versus another, a reference category of the dependent variable needs to be selected for their calculation. Take-up, irrespective of the developer, appears more likely when the drug candidate has scientific references to external research, be it basic or applied. Applied internal SNPRs only show a significant coefficient for take-up by other firms when compared to no take-up, although results in the third column indicate no significant difference when comparing own versus other take-up. Basic internal SNPRs show only a significant influence on take-up by the patent owner, with also a significant difference between own and other take-up as shown in the third column.

Among the control variables, it is salient that a firm's focus on basic research is positively associated with take-up by the patent owner. This may indicate that a strong basic science orientation at the firm increases the willingness or capability for taking up risky projects in-house. The size of the firm's pipeline only influences the decision to internalize development: drug candidates of experienced firms are more likely to be developed in-house, though this effect diminishes when the pipeline is well filled. The share of the patent owner within the respective market for technology strongly influences the take-up decision: a positive relationship with take-up by the patent owner and a negative relationship with take-up by other

[9] Table A2 in appendix presents the average marginal effects of these regressions.

firms. This finding supports the idea that firms select the most capable developer to take on clinical trials.[10]

1.8 SUPPLEMENTARY ANALYSES

1.8.1 Average marginal effects

As the coefficients of Logit regression models cannot be interpreted as marginal effects, Table A1 and A2 present the average marginal effects across observations for the main regression models of respectively Table 4 and Table 5.

1.8.2 Alternative identification of take-up into a development project

When considering take-up into development as an investment decision, it can be argued that firms already decide when starting pre-clinical research on a specific drug candidate. While the project data does mention preclinical research, there are multiple reasons for my choice to define take-up as the start of clinical research. Frist, the internal validity on the reported pre-clinical research is questionable. Of the 698 patents that are mentioned within the project data (at least registering pre-clinical research), 449 mention discovery or preclinical research. However, 247 patents that mention clinical research or more never refer to preclinical research or discovery. Hence, it seems that the dataset does not consistently identify preclinical research. The preclinical stage is less regulated diminishing the need for reporting and making it more difficult to pin down within a project-level dataset. Moreover, previous research has indicated that it is difficult to assign preclinical research, let alone discovery, to one single drug (DiMasi, Hansen, and Grabowski 2003). It is unclear to me whether the start-up of preclinical research indicates that the firm has interest in the research domain or that it has a specific promising discovery. Third, the study setup only entails selection on pre-clinical research as it investigates patented drug candidates. Because patents inquire an indication of usefulness and entail undeniable application costs, they are typically applied for after some pre-clinical research and before clinical trials (Campbell, 2005; PPD, 2017). Fourth, due to the extent costs incurred during clinical research (DiMasi, Hansen, and Grabowski 2003). The take-up decision obtains more attention within the clinical research debate than the decision on advancing fundamental research into pre-clinical investigation. Finally, the regression results when take-up is defined

[10] When comparing patent owners and their selected external development partner, the data further supports this finding as the partner significantly outperforms the patent owner in size of the pipeline and finalized projects in the last ten years.

as reaching preclinical research, which are presented in Table A3 of the appendix, are similar to the main findings. We observe 698 drug candidates which are taken up into pre-clinical research of which 541 clinical trials.

1.8.3 Development success

In this study, we analysed the take-up decision of drug candidates into a development project. When making this decision, a firm will consider both the expected probability of successful completion of the clinical trials and the expected value of the marketed drug. However, firms can misjudge this expectation, especially the probability on successful development, as it can be subject to new scientific breakthroughs, actions of competitors, and unknown elements of the disease, compound or effects on the human body. In this supplementary analysis, we want to explore how the scientific background of the patented drug candidate can still explain some of the variance in development success of selected candidates.

There are four possible outcomes after drug candidates are taken-up into clinical trials: (1) failed take-up by other firms, (2) successful take-up by other firms, (3) failed take-up by patent owner, (4) successful take-up by patent owner[11]. We perform two logit regressions respectively on those drug candidates taken up by another firm (279 observations) and taken up by the patent owner (305 observations)[12]. Additional to the patent-level controls, we include controls of the development project and of the firm leading clinical development: the number of indications of the project, the activity within the ATC classes of the project[13], the development experience of the developing company, the past success rate in development of the developing company and the pipeline of the developing company. Within Table A4, the second column shows the results for successful development when the project was taken up by other firms, in respect to failure of these projects. The third column presents the results of success for projects taken up by the patent owner. When examining our variables of interest, the coefficients present barely any

[11] Success is defined as reaching pre-registration at a drug agency or in other words successful completion of clinical phase III. However, the results are similar when applying alternative thresholds for success: successful completion of clinical phase I, successful completion of clinical phase II or drug approval. However these findings show limited significance indicating the importance of the take-up decision as a pre-selection of successful projects.

[12] In the analysis of the take-up decision, the level of observation is the patent. However, 62 patents are linked to more than one project. In this analysis, all projects are considered as a separate observation with their own project and developer level controls resulting in multiple observations for one patented drug candidate.

[13] This variable is measured by the average weighted number of projects started in the project's ATC classes in the past four years.

significance. Only the presence of applied external SNPRs appears to increase the success rate of projects taken up by the patent owner. For this success analysis, we use preregistration to identify development success as it is the point upon which doctors can write prescriptions for the drug and because most of the preregistered drugs also obtain approval at a federal drug office[14]. Alternatively, approval can be used as cut-off point giving highly similar results as shown in Table A5.

These supplementary analyses do not provide support for an increased advantage in development of the patent owner when the patent cites internal basic research. Moreover, they do not signal that drug candidates with basic internal SNPRs are less successful as their take-up was based on irrational grounds like over-commitment, or not-sold-here syndrome. However, these results should be interpreted with caution. First, the results are of low trustworthiness as the samples sizes are small. Second, the development process is highly complex and long, making it less likely that one can explain ongoing development success by elements observed at the time of application of the patent. Multiple events may occur down the road both under control of the developing firm (mergers, strategic preference for other drug candidate, financial issues,…) and out of its control (changed legislation, new scientific discoveries, new drugs on the market,…). Moreover, these regressions examine the success rate in a certain phase *conditional* on take-up into clinical trials. As the take-up decision is characterized by the highest fall-out rate, there is a major selection bias, which these models do not into account. As also indicated by Ceccagnoli and Jiang (2013), the take-up decision itself will already depend on firms' expectation of safety, efficacy and market value of the drug candidate.

1.8.4 Take-up bias for patents linked to internal research

Within the theoretical argumentation, we remarked that the take-up decision for drug candidates with an internal basic research background might suffer from irrational consideration. This reasoning implies that our findings may reflect 'endogeneity' of internal SNPRs or i.e. pre-selection of drug candidates for take-up into development. If firms aim to be a significant player in certain commercial markets, their science strategy will most likely involve the set-up of discovery or preclinical research within the relevant scientific fields. Hence, firms may have a

[14] Only 4 of the 129 pre-registered patents taken up by another firm are not registered and 1 of the 93 pre-registered patents taken up by the patent owner are not registered.

positive bias towards taking up patents that are based on own science compared to those patents that only cite external science.

To examine this explanation in more detail, we performed two robustness checks. First, we included an additional dummy variable within Model 4 indicating patents that only cite external science. The results presented in Table A6 show that these patents are less likely to be taken-up by the patent owner, though the effect is insignificant. Additionally, the effect of basic internal SNPRs becomes insignificant. However, as the majority (97%) of patents that are based on internal science cite basic internal research, the collinearity between 'only external SNPRs' and 'basic internal SNPRs' is too high (98.3) to make statements on their individual effects.

Second, we examined whether drug development projects that specifically include a discovery or preclinical research step within our database, are more likely to be developed by the patent owner. These projects were initiated within the scientific strategy of the firm with early expectations on commercial value, though, there output, the patented drug candidate, may not be directly linked to the results of this internal basic science. These patents, that fit the firm's scientific strategy, represent 0.87 % of the drug candidates that are not taken up, 54.34 % of those that are taken up, 48.08 % of those taken-up in-house and 61.42% of those taken up by other firms. The model in Table A7 includes a dummy variable representing patents with a stated discovery or preclinical phase. The development of these drug candidates appears more likely overall. To our surprise, there is no significant difference to be noted in the effect of these patents on take-up by the patent owner or by other firms.

Hence, while it may be argued from a not-sold-here or commitment argument that internal scientific projects are more likely to continue development internally, this is not confirmed by the supplementary analysis. A possible interpretation is that while a scientific strategy is set up to generate profits in certain markets, this can be realized by both own development and licensing revenues (take-up by others). A second explanation may rest in the internal organization and management of the take-up decision (Lichtenthaler 2011). Due to the highly difficult decision of take-up and the sensitivity of project decisions to irrational feelings (Granovetter, 2013; Antons and Piller, *2015; Peck et al.*, 2015; Eliëns *et al.*, 2018), firms may have organized discovery and development within individual units. More so, companies may place the responsibility of the take-up decision in independent units, like the financial analysts of the pharmaceutical development and manufacturing sciences department of J&J. Also during a Tufts CSDD R&D Management gathering in 2012 managers put forward the need to avoid existing silos and encourage rapid handoffs between teams when organizing the transition from

nonclinical to early clinical development (Tufts Center for the Study of Drug Development 2012). Finally, firms may apply different strategies towards the market for technology depending on the technological field of the drug candidate, for example avoiding licensing in core fields (Ceccagnoli and Jiang 2013). Even though, we do not observe aggregated evidence of the continuation of internal research projects into development based on irrational considerations, it may be present in particular technological fields.

1.8.5 The take-up decision of successful/unsuccessful decision makers

One limitation of our data is that we do not know whether firms made the 'best' decision. We have assumed that firms, which aim for profit maximization, will think rationally and will select the option with the highest expected value under a certain probability of successful development. However, as corporate decisions are made by individuals or teams, we know that they are not always rational (Granovetter, 2013; Antons and Piller, 2015; Peck *et al*, 2015). Firms that are able to detain these irrational influences from their take-up decision (rational decision makers), should obtain higher success rates across their patented drug candidates. Hence, by comparing the take-up decision of highly successful firms to that of the others, we hope to obtain some insight on how the scientific background of a drug candidate relates to the 'best' development decision.

Tables A8 and A9 presents the regression results of a split sample analysis based on the success rate of the patent owner. More precisely, we identified successful decision makers as firms of which the taken-up drug candidates have a success rate above the observed mean: having a probability on reaching pre-registration at a drug administration office above 26.2%. The remaining firms in the sample are considered unsuccessful decision makers. The two tables show similar results for 'basic internal SNPRs' indicating that also firms that are assumed to make more rational decisions on take-up prefer to develop these drug candidates in-house. The figures do show that the unsuccessful firms are mainly driven by external applied knowledge to make their take-up decision. This may indicate that these firms mainly work on exploitative innovation, further refining and repurposing existing drugs. However, more detailed investigation would be needed to verify this interpretation.

Sample restriction to patented drug candidates applied for at the USPTO

Patent offices maintain diverging rules and expectation on the inclusion of SNPRs. For example, at the USPTO the applicant needs to submit a comprehensive set of prior art, while at the EPO the patent examiner is known to perform the most extensive search of prior art. Within

the data it is for instance noticeable that patents applied for at the USPTO are significantly more likely to cite scientific research than the other patents (t-test: 61.14, p 0.000). Nonetheless, when limiting our sample to patent families with at least one application at the USPTO, the results remain similar, as shown in Table A10. Only external basic research references appear to have a limited significance on take-up overall while its coefficient was highly significant for the full sample. This may be due to the overwhelming presence of these references among the USPTO applications: 72 percent versus 28 percent for non-USPTO applications (t-test: 58.31, p 0.000).

1.8.6 Alternative sample of patents listing a therapeutic use

As mentioned in the data section, the patent sample of this study has been selected to represent pharmaceutical compounds that are regarded as the first patents within the drug development process. However, it is highly difficult to verify the accuracy of our sample. In the future, we may attempt to determine the presence of a Markush structure on the patent (Jayaraj and Gittelman 2018). While this method and also the sample selection (IPC A61 and (C07 or C12)) focusses mostly on compounds, an alternative sample could focus on the pharmaceutical use of the patent represented by IPC class A61P 'Specific Therapeutic Activity of Chemical Compounds or Medicinal Preparations'. This results in a smaller sample of 16,144 patents, which overlaps only partially with the original sample (13,385 patents). The set of patents taken-up into clinical development within this sample is higher: 570 instead of 541. Table A11 presents the results of the alternative sample, which are similar to the main findings of the paper, except that the coefficient for internal applied SNPRs is also significant for explaining take-up by the patent owner. Moreover, when looking only at the overlapping patents (A61P and (C07 or C12)) patents also shows robustness of our findings as indicated by Table A12. This highly restrictive sample only takes into account the patents that present a pharmaceutical use and protect chemical or biological compounds. This avoids follow-up patents that protect new usages of existing compounds or drugs. This highly selective sample consists out of 13,385 patents of which 231 are developed by the patent owner and 280 by only other firms. Hence, there is an observed probability on take-up of 3.8 percent (511/13,385).

1.8.7 IIA assumption of multinomial logit

In this chapter, we applied multinomial logit regression to analyse the take-up with several possible outcomes: no take-up, own tale-up, take-up by other firms. However, the multinomial regression makes the strong assumption of independence from irrelevant alternatives (IIA assumption). This assumption implies that the ratio of the probabilities of choosing any two

alternatives is independent of the attributes of any other alternative in the choice set (Hausman and Mcfadden 1984).

Empirically, the IIA assumption can be verified by the Hausman test. When applied to our data, the test shows a negative CHI statistic. While this indicates that the null hypothesis cannot be rejected by the available data, it also clearly shows that the data is too limited to properly perform the test. However, further support can be found in Table A13 that shows that separate logit regressions for take-up by other firms and take-up by the patent owner in respect to no take-up result in similar findings to the multinomial logit regression.

Theoretically, the IIA assumption would indicate that the owner of a patented drug candidate would make the same decision on internal take-up when there is no other firm available that could take-up the specific drug candidate into development then when there would be an outside development option. Similarly, a drug candidate would be as likely to be undeveloped when it had the option to be developed at its patent owner than when it would not have this option. Crucial for the validity of the IIA assumption in our setting is that one patented drug candidate can be taken up in multiple development projects. Hence, a single drug candidate can be taken-up by a patent owner while also developed by another firm in another project, and vice versa. Computationally, the outcome of own take-up just entails that the drug candidate starts clinical trials in at least one development project of the patent owner. Only when the patent owner is not involved in a development project of the candidate the outcome will signal development by others.

Alternative to the take-up decision could also be modelled as a nested logit regression, first explaining whether someone will take-up the drug candidate and then regressing whether the patent owner is one of the companies taking up the development. This would allow correlation between take-up by others and take-up by the patent owner. Moreover, the first stage regression would signal the market value of the drug candidate, while the second stage regression would indicate its intra-firm evaluation. I find this a highly interesting point of view and would like to explore it in more detail in the future.

1.9 DISCUSSION

Our findings indicate that both the presence of basic and applied scientific research within the list of SNPRs has a positive effect on the probability of take-up in clinical drug development. Moreover, basic and applied scientific research appear of equal importance for the take-up

decision. We posit that drug candidates with an origin in basic science have a higher probability to be radical (Della Malva et al., 2015) increasing their expected market value. Basic scientific insights may also enable firms to better select fruitful clinical trial setups (Fabrizio 2009; Fleming and Sorenson 2004), increasing their expected success rate in development. Applied scientific origins of drug candidates may bring increased confidence in successful development of the drug, as applied research provides knowledge on crucial clinical research techniques and is likely to inform firms on viable development approaches.

Basic research performed internally is the only type of scientific background that is not related to development by other firms. We posit that this is because the tacit knowledge derived from basic research is not easily codified and transferred to external parties (Lynne G. Zucker, Darby, and Armstrong 2002a), hence providing not much relevant information on the success rate or better execution of the project to other firms. The focal firm as licensor may also experience strategic and competitive risk when licensing patents based on fundamental scientific research performed in-house, with adequate transfer to the licensee risking (unintended) spill-overs of core competences (Cockburn 2007; Klueter, Monteiro, and Dunlap 2017). However, the presence of basic internal SNPRs is associated with a higher probability of in-house development. Firms' decision process may be influenced by the not-sold-here syndrome (Antons and Piller 2015; Natalicchio, Messeni Petruzzelli, and Garavelli 2014), optimism bias (Peck *et al.*, 2015) and over-commitment (Eliëns et al. 2018) stimulating internal development even though the drug candidate may be of limited value, development may be difficult or the focal firm is not the most suited developer. Alternatively, internal development may be the rational choice because the embedder of relevant fundamental knowledge may be the most capable developer: better able to select the most effective set-up of clinical trials, correctly interpret test results and quickly react on expected outcomes. The main regression analysis of this study is unable to distinguish between these two mechanisms. The supplementary analyses showed that development success of internal development projects is not significantly influenced by basic internal SNPRs, acknowledged pre-clinical research had an overall stimulating impact on take-up and the decision process of successful firms is similarly effected by basic internal SNPRs as unsuccessful firms. Hence, while not providing supportive evidence for the mechanism of enhanced development capability, the supplementary analyses reject the possibility of continuation of research projects based on irrational considerations.

When the drug candidate, however, refers to in-house applied research, also other firms appear more likely to take on development. Thanks to the codified and practical nature of applied

scientific understanding (Gittelman 2016), the insights of these references may be transferred between the patent owner and a partner developer without significant problems. Moreover, knowledge from prior clinical trials may enable the formation of more precise expectations on the value of the drug candidate reducing information asymmetry and the risk on buying lemons (Pisano 1997a). In other words, the presence of applied publications on the internal research of the patent owner may inform another firm as it does the patent owner, leading to a similar effect on the take-up rate.

1.10 CONCLUSION

Driven by the discussions on scientific productivity and in-house scientific research within the pharmaceutical industry, we examined whether the scientific background of drug candidates influences the probability of take-up in clinical development projects. We distinguished development projects initiated by the owner of the patented drug candidate and projects initiated by other firms. Moreover, we categorized the scientific background of the patented drug candidate according to two dimensions: basic/applied and internally/externally developed.

Our empirical study investigated 18,363 drug candidates for which 131 pharma and biotech companies applied patent protection during 1995-2002 and for which development activity was observed until 2013. The results show that all scientific references of these patents have a positive and significant effect on the probability of take-up in a development project. While previous studies have indicated the importance of basic research for technological innovation (Adams, 1990; Sorenson, Rivkin and Fleming, 2006; Della Malva *et al*., 2015; Arora, Belenzon and Sheer, 2017), our study indicates that, besides basic, also applied research is strongly valued by firm's deciding on their development investments.

Our study emphasizes the importance of in-house basic research to enlarge firms' drug pipeline. Involvement in the basic research underlying the patented drug candidates leads to clinical development by the patent owner. Two possible explanation have come forward. First, embedding fundamental background knowledge on the drug candidate may provide a competitive advantage of the patent owner in the development process. Second, irrational considerations like over-commitment, over-optimism and not-sold-here syndrome may stimulate firms to extend investments from research to development. As the supplementary analyses contradict the second explanation, we posit that the main mechanism for enhanced in-house development of drug candidates with basic internal SNPRs is the enhanced developer

capabilities of the patent owner thanks to embedded fundamental knowledge. Other firms' take-up (e.g. through licensing) is not influenced by the presence of internal basic research insights, as the knowledge may be too tacit and embedded to effectively transfer and evaluate. In contrast, applied internal research appears to influence take-up by the patent owner and by other firms in a similar manner, as the related codified knowledge may more easily be transferred and helps in overcoming the lemons problem.

Besides our contribution to the examination of the relationship between scientific research and development, we contribute to two important discussions in the literature. First, we provide further evidence for the innovative benefits of in-house scientific research. Our findings suggest that it may not only increase absorptive capacity (Arora, Belenzon, and Sheer 2017; Arora, Belenzon, and Patacconi 2015; Stijn Kelchtermans, Lete, and Belderbos 2013), but may also stimulate innovation by spurring investments in internal development. Second, by separating development by the patent owner and other firms, our findings reflect on the potential consequences of the structure of the pharmaceutical industry. While our results are not conclusive, they support the idea of complementarity between scientific and commercial activities as suggested by the opponents of the disaggregation (Cockburn 2007; Arora, Belenzon, and Sheer 2017; Arora, Belenzon, and Patacconi 2015). In-house development appears more common for firms with stronger basic research focus and patents with internal basic SNPRs. In other words, firms expect to benefit from their basic scientific research efforts, directly or indirectly. This expectation may spur from irrational considerations stimulating firms to extend investments from research to development. However, we find stronger indications that it is based on stronger internal development capabilities thanks to embedded fundamental knowledge. This would imply that vertical integration facilitates the link between basic research and clinical development.

Our study has its limitations. First, we only examine patent and firm characteristics at the moment of application. However, firms can take actions after patent application that may benefit the development process, like collaboration (Crispeels, Willems, and Scheerlinck 2018). Second, the project database refers to the lead company as responsible party of the clinical trials, though it is unclear at which point during the development process this responsibility is taken up. Third, we approximate the scientific background of a drug candidate by the scientific references of its protecting patent. Though, as for every approximation, comments can be made on the coverage of this measurement (e.g. Callaert *et al.* 2013). The patent owner or the project leader may not know some references. Moreover, advances in science after the patent

application may also contribute to the take-up or success of a drug development project. Fourth, we assume that all patents in our sample have the potential to serve as drug candidate. However, also strategic grounds can underlie patent applications (Thumm 2004; Blind, Cremers, and Mueller 2009). While we aim to mitigate this issue by focusing on patents covering new compounds, this concern cannot be completely eliminated. Finally, we do not take into account the different procedures and practices in the listing of SNPRs across patent offices while we examine patents applied for at all patent offices covered by PATSTAT. However, restricting our analysis to patents applied at the USPTO does not affect the main findings of our study.

As the investigation of project data is an upcoming stream in the literature, still many research avenues are open for future research. We would welcome future research on three topics that appear to influence our results. The transition of drug development processes between firms is a first topic that deserves additional attention of scholars: at which stage in the development process are drug candidates transferred to other companies, are their differences to be noted between transitions by outsourcing or licensing. Second, scholars can investigate in more detail who is taking up the drug candidate into a development project. In particular, they may examine the importance of knowledge of the underlying science (e.g. by examining prior publications of the firms) besides firms' development capabilities.

1.11 APPENDIX

Table 1.A1. Logit regression on the probability of take-up into clinical trials: A.M.E.

	Model 1 Controls	**Model 2 Basic-Applied**	**Model 3 Internal-External**
	A.M.E. (S.E.)	A.M.E. (S. E.)	A.M.E. (S.E.)
D(basic SNPRs)		0.015756***	
		(0.003182)	
D(applied SNPRs)		0.017366***	
		(0.002848)	
D(basic internal SNPRs)			0.008307***
			(0.002937)
D(applied internal SNPRs)			0.015012**
			(0.006408)
D(basic external SNPRs)			0.012592***
			(0.003155)
D(applied external SNPRs)			0.013498***
			(0.002987)
No. of owners	-0.004508	-0.003655	-0.004304
	(0.004083)	(0.003975)	(0.003962)
Scope: # IPC6 classes	0.002186***	0.002019***	0.002017***
	(0.000218)	(0.000217)	(0.000221)
No. of inventors	0.001494***	0.001576***	0.001577***
	(0.000294)	(0.000299)	(0.000299)
Log(R&D $_{\text{in million \$, in t}}$)	-0.002732***	-0.002424***	-0.002461***
	(0.000616)	(0.000630)	(0.000624)
Science intensity $_{\text{from t-4 to t-1}}$	0.000967	0.000502	0.000115
	(0.000776)	(0.000867)	(0.000941)
Basic research focus $_{\text{from t-4 to t-1}}$	0.010508	0.013776	0.012241
	(0.011389)	(0.011369)	(0.011653)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.009391**	0.007934*	0.008085*
	(0.004412)	(0.004501)	(0.004531)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	-0.000643	-0.000549	-0.000548
	(0.000491)	(0.000476)	(0.000479)
Size of market for technology	-0.000023	-0.000024	-0.000025
	(0.000030)	(0.000030)	(0.000031)
Share in market for technology	0.208494***	0.225625***	0.215122***
	(0.060763)	(0.061357)	(0.061723)
Year dummies	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES
Biotech compounds dummy	YES	YES	YES
Chemical compounds dummy	YES	YES	YES

* p<0.10, ** p<0.05, *** p<0.01.. Robust standard errors are reported in brackets.

Table 1.A2. Multinomial Logit regression results explaining drug candidates not being taken up (1), taken up by other firms (2) and taken up by patent owner (3): A.M.E.

	Model 4		
	1. Not taken up	**2.Taken up by other firms**	**3. Taken up by owner**
	A.M.E. (S.E.)	A.M.E. (S. E.)	A.M.E. (S.E.)
D(basic internal SNPRs)	-0.008245***	0.000308	0.007937***
	(0.002974)	(0.002112)	(0.002128)
D(applied internal SNPRs)	-0.016068**	0.008968**	0.007100
	(0.006445)	(0.004267)	(0.004752)
D(basic external SNPRs)	-0.012468***	0.004645**	0.007823***
	(0.003169)	(0.002163)	(0.002421)
D(applied external SNPRs)	-0.013008***	0.007540***	0.005468**
	(0.002983)	(0.002096)	(0.002182)
No. of owners	0.003050	-0.001439	-0.001611
	(0.003966)	(0.002637)	(0.003084)
Scope: # IPC6 classes	-0.001966***	0.000996***	0.000970***
	(0.000224)	(0.000144)	(0.000153)
No. of inventors	-0.001587***	0.000762***	0.000825***
	(0.000304)	(0.000185)	(0.000244)
Log(R&D $_{\text{in million \$, in t}}$)	0.001882***	-0.000736*	-0.001146**
	(0.000679)	(0.000430)	(0.000539)
Science intensity $_{\text{from t-4 to t-1}}$	0.000470	0.000102	-0.000572
	(0.000995)	(0.000742)	(0.000697)
Basic research focus $_{\text{from t-4 to t-1}}$	-0.028886**	-0.006998	0.035884***
	(0.013777)	(0.005663)	(0.012934)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	-0.024924***	0.001811	0.023113***
	(0.006003)	(0.002020)	(0.005802)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	0.005588**	0.000081	-0.005669**
	(0.002298)	(0.000138)	(0.002366)
Size of market for technology	0.000020	-0.000010	-0.000009
	(0.000031)	(0.000021)	(0.000023)
Share in market for technology	-0.014643	-0.263479***	0.278122***
	(0.101429)	(0.095852)	(0.041299)
Year dummies	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES
Biotech compounds dummy	YES	YES	YES
Chemical compounds dummy	YES	YES	YES

* p<0.10, ** p<0.05, *** p<0.01.. Robust standard errors are reported in brackets.

Table 1.A3. Logit and Multinomial Logit regression results. Take-up defined as the start of pre-clinical research

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.22**	-0.05	0.46***	0.51***
	(0.10)	(0.15)	(0.13)	(0.19)
D(applied internal SNPRs)	0.41*	0.61**	0.30	-0.30
	(0.23)	(0.30)	(0.32)	(0.42)
D(basic external SNPRs)	0.38***	0.37***	0.38***	0.00
	(0.10)	(0.14)	(0.13)	(0.19)
D(applied external SNPRs)	0.46***	0.53***	0.37***	-0.15
	(0.10)	(0.14)	(0.13)	(0.19)
No. of owners	-0.13	-0.20	0.03	0.22
	(0.13)	(0.18)	(0.17)	(0.25)
Scope: # IPC6 classes	0.07***	0.07***	0.07***	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.07***	0.07***	0.07***	-0.00
	(0.01)	(0.01)	(0.01)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.11***	-0.07***	-0.11***	-0.04
	(0.02)	(0.03)	(0.03)	(0.04)
Science intensity $_{\text{from t-4 to t-1}}$	0.02	0.03	-0.04	-0.07
	(0.03)	(0.04)	(0.04)	(0.06)
Basic research focus $_{\text{from t-4 to t-1}}$	0.19	-0.47	1.52**	1.98**
	(0.36)	(0.39)	(0.69)	(0.78)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.39***	0.15	1.58***	1.42***
	(0.14)	(0.14)	(0.26)	(0.29)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	-0.03*	-0.01	-0.31***	-0.30***
	(0.02)	(0.01)	(0.07)	(0.07)
Size of market for technology	-0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	5.57***	-28.67***	16.18***	44.85***
	(2.09)	(8.39)	(2.25)	(8.67)
Year dummies	YES	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES	YES
Biotech compounds dummy	YES	YES	YES	YES
Chemical compounds dummy	YES	YES	YES	YES
Constant	-3.07***	-2.94***	-5.51***	-2.57***
	(0.44)	(0.51)	(0.81)	(0.94)
N	18363.00	18363.00		18363.00
LL	-2642.89	-3039.19		-3039.19

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A4. Logit regressions explaining development success (pre-registration), by other firm and by patent owner

Regression outcome	**Successful take-up by other firms**	**Successful take-up by owner**
	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.45	0.18
	(0.40)	(0.36)
D(applied internal SNPRs)	0.49	0.27
	(0.79)	(0.73)
D(basic external SNPRs)	-0.06)	0.10
	(0.42)	(0.51)
D(applied external SNPRs)	0.25	0.88**
	(0.42)	(0.41)
No. of owners	0.11	-1.10
	(0.58)	(0.76)
Scope: # IPC6 classes	0.03	-0.02
	(0.02)	(0.03)
No. of inventors	-0.08	-0.01
	(0.06)	(0.05)
No. of indications of project	0.06	0.27***
	(0.08)	(0.06)
Activity in project's ATC class $_{\text{in t-4 to t-1}}$	-0.08***	0.23**
	(0.03)	(0.10)
Development experience $_{\text{t-10 to t-1}}$	-0.87	-0.65
	(1.23)	(0.69)
Past success rate $_{\text{t-10 to t-1}}$	0.53	-0.64
	(5.50)	(2.52)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	-0.00	0.00
	(0.00)	(0.01)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$^2	-0.00	-0.00
	(0.00)	(0.00)
Year dummies	YES	YES
22 Therapeutic field dummies	YES	YES
Biotech compounds dummy	YES	YES
Chemical compounds dummy	YES	YES
Constant	0.40	-5.00***
	(1.27)	(1.45)
N	279.00	305.00
LL	-136.53	-124.23

* p<0.10, ** p<0.05, *** p<0.01. Regressions include year dummies, therapeutic field dummies, biotech compounds dummy, chemical compounds dummy and medical technology dummy. Robust standard errors are reported in brackets.

Table 1.A5. Logit regressions explaining development success, measured by registration (rather than pre-registration), by other firm and by patent owner

Regression outcome	**Successful take-up by other firm**	**Successful take-up by owner**
	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.35	0.25
	(0.42)	(0.35)
D(applied internal SNPRs)	0.48	0.28
	(0.80)	(0.75)
D(basic external SNPRs)	0.09)	-0.01
	(0.45)	(0.51)
D(applied external SNPRs)	0.32	0.93**
	(0.45)	(0.42)
No. of owners	0.21	-1.07
	(0.59)	(0.76)
Scope: # IPC6 classes	0.02	-0.02
	(0.02)	(0.03)
No. of inventors	-0.06	-0.01
	(0.06)	(0.05)
No. of indications of project	0.08	0.27***
	(0.08)	(0.06)
Activity in project's ATC class $_{\text{in t-4 to t-1}}$	-0.07**	0.26***
	(0.03)	(0.10)
Development experience $_{\text{t-10 to t-1}}$	-0.90	-0.69
	(1.22)	(0.70)
Past success rate $_{\text{t-10 to t-1}}$	0.52	-0.72
	(5.62)	(2.54)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	-0.00	0.00
	(0.00)	(0.01)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	-0.00	-0.00
	(0.00)	(0.00)
Year dummies	Yes	Yes
22 Therapeutic field dummies	Yes	Yes
Biotech compounds dummy	Yes	Yes
Chemical compounds dummy	Yes	Yes
Constant	-0.43	-4.92***
	(1.28)	(1.45)
N	279.00	305.00
LL	-133.00	-122.79

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A6. Logit and Multinomial Logit regression results, controlling for only external SNPRs

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.62	0.90	0.40	-0.50
	(0.51)	(0.68)	(0.66)	(0.85)
D(applied internal SNPRs)	0.68***	0.93***	0.52*	-0.41
	(0.21)	(0.28)	(0.30)	(0.38)
D(basic external SNPRs)	0.45***	0.42***	0.46***	0.05
	(0.09)	(0.13)	(0.13)	(0.18)
D(applied external SNPRs)	0.51***	0.66***	0.31**	-0.35**
	(0.09)	(0.12)	(0.13)	(0.17)
D(only external SNPRs)	0.32	0.86	-0.15	-1.01
	(0.52)	(0.69)	(0.67)	(0.87)
No. of owners	-0.16	-0.21	-0.01	0.21
	(0.12)	(0.17)	(0.17)	(0.23)
Scope: # IPC6 classes	0.08***	0.08***	0.08***	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.06***	0.06***	0.05***	-0.02
	(0.01)	(0.01)	(0.01)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.10***	-0.05**	-0.11***	-0.05
	(0.02)	(0.03)	(0.03)	(0.04)
Science intensity $_{\text{from t-4 to t-1}}$	0.00	-0.01	-0.03	-0.02
	(0.03)	(0.05)	(0.04)	(0.06)
Basic research focus $_{\text{from t-4 to t-1}}$	0.50	-0.31	2.49***	2.80***
	(0.35)	(0.34)	(0.76)	(0.83)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.41***	0.26*	2.01***	1.75***
	(0.13)	(0.14)	(0.32)	(0.35)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$^2	-0.03**	-0.01	-0.50***	-0.49***
	(0.01)	(0.01)	(0.14)	(0.14)
Size of market for technology	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	6.55***	-19.74***	17.39***	37.13***
	(1.92)	(5.81)	(2.26)	(6.20)
Year dummies	YES	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES	YES
Biotech compounds dummy	YES	YES	YES	YES
Chemical compounds dummy	YES	YES	YES	YES
Constant	-3.46***	-3.83***	-5.99***	-2.15*
	(0.65)	(0.82)	(1.07)	(1.29)
LL	-2978.07	-3433.25		-3433.25

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A7. Logit and Multinomial Logit regression results, controlling for stated discovery and preclinical research fitting the firm's scientific strategy

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.28**	0.07	0.43***	0.36**
	(0.12)	(0.15)	(0.14)	(0.18)
D(applied internal SNPRs)	1.00***	1.04***	1.05***	0.02
	(0.20)	(0.25)	(0.26)	(0.31)
D(basic external SNPRs)	0.36***	0.28*	0.39***	0.11
	(0.12)	(0.14)	(0.15)	(0.18)
D(applied external SNPRs)	0.51***	0.68***	0.27*	-0.41**
	(0.11)	(0.15)	(0.15)	(0.18)
D(fitting scientific strategy)	4.92***	5.04***	4.83***	-0.21
	(0.11)	(0.14)	(0.14)	(0.18)
No. of owners	-0.22	-0.30*	-0.04	0.26
	(0.15)	(0.18)	(0.19)	(0.21)
Scope: # IPC6 classes	0.06***	0.06***	0.06***	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.02	0.02	0.02	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.03	0.05	-0.08***	-0.13***
	(0.03)	(0.03)	(0.03)	(0.04)
Science intensity $_{\text{from t-4 to t-1}}$	-0.03	-0.07	-0.06	0.01
	(0.04)	(0.06)	(0.04)	(0.06)
Basic research focus $_{\text{from t-4 to t-1}}$	0.24	-0.25	1.51***	1.76***
	(0.34)	(0.38)	(0.57)	(0.60)
Size of Pipeline $_t$/R&D$_{\text{in million \$, t}}$	0.00***	-0.00***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Size of Pipeline $_t$ /R&D$_{\text{in million \$, t}}$^2	-1.74**	-4.40***	0.23	4.64***
	(0.77)	(1.35)	(0.75)	(1.36)
Size of market for technology	0.19	-0.04	2.74***	2.77***
	(0.17)	(0.11)	(0.41)	(0.43)
Share in market for technology	-0.02	0.00	-0.83***	-0.83***
	(0.01)	(0.00)	(0.23)	(0.23)
Year dummies	YES	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES	YES
Biotech compounds dummy	YES	YES	YES	YES
Chemical compounds dummy	YES	YES	YES	YES
Constant	-4.86***	-4.66***	-7.78***	-3.12***
	(0.40)	(0.45)	(0.66)	(0.73)
LL	-2202.35	-2712.02		-2712.02

* p<0.10, ** p<0.05, *** p<0.01. Regressions include year dummies, therapeutic field dummies, biotech compounds dummy, chemical compounds dummy and medical technology dummy. Robust standard errors are reported in brackets.

Table 1.A8. Multinomial Logit regression results for successful firms

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.30**	0.04	0.48***	0.43
	(0.14)	(0.22)	(0.18)	(0.28)
D(applied internal SNPRs)	0.74**	0.97**	0.67*	-0.30
	(0.31)	(0.46)	(0.40)	(0.59)
D(basic external SNPRs)	0.79***	0.85***	0.76***	-0.09
	(0.16)	(0.26)	(0.20)	(0.33)
D(applied external SNPRs)	0.41***	0.56**	0.27	-0.29
	(0.15)	(0.23)	(0.19)	(0.30)
No. of owners	-0.39*	-0.38	-0.42	-0.04
	(0.22)	(0.37)	(0.27)	(0.46)
Scope: # IPC6 classes	0.07***	0.08***	0.07***	-0.01
	(0.01)	(0.01)	(0.01)	(0.02)
No. of inventors	0.07***	0.05**	0.08***	0.03
	(0.02)	(0.02)	(0.02)	(0.03)
Log(R&D $_{\text{in million \$, in t}}$)	-0.12***	-0.11*	-0.10**	0.00
	(0.04)	(0.06)	(0.05)	(0.07)
Science intensity $_{\text{from t-4 to t-1}}$	0.02	0.04	-0.07	-0.11
	(0.08)	(0.12)	(0.11)	(0.16)
Basic research focus $_{\text{from t-4 to t-1}}$	-0.23	-1.15*	1.71**	2.87***
	(0.53)	(0.62)	(0.83)	(1.01)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.18	0.39	3.62**	3.24**
	(0.22)	(0.29)	(1.42)	(1.45)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$^2	-0.01	-0.02	-3.20**	-3.18**
	(0.02)	(0.02)	(1.53)	(1.53)
Size of market for technology	-0.00	0.00	-0.00	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	8.36***	-6.49	11.31***	17.80**
	(2.69)	(6.32)	(3.46)	(7.11)
Year dummies	YES	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES	YES
Biotech compounds dummy	YES	YES	YES	YES
Chemical compounds dummy	YES	YES	YES	YES
Constant	-2.88***	-3.02***	-5.56***	-2.54**
	(0.65)	(0.85)	(0.99)	(1.29)
LL	-1210.69	-1370.25		-1370.25
N	10,491	10,491		10,491

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A9. Multinomial Logit regression results for unsuccessful firms

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.34*	0.16	0.54**	0.38
	(0.18)	(0.24)	(0.26)	(0.34)
D(applied internal SNPRs)	0.45	0.42	0.35	-0.07
	(0.38)	(0.46)	(0.62)	(0.74)
D(basic external SNPRs)	0.08	0.03	0.19	0.16
	(0.17)	(0.22)	(0.28)	(0.35)
D(applied external SNPRs)	0.54***	0.56***	0.59**	0.03
	(0.16)	(0.22)	(0.25)	(0.32)
No. of owners	0.24	0.07	0.64**	0.57
	(0.19)	(0.24)	(0.31)	(0.39)
Scope: # IPC6 classes	0.07***	0.07***	0.08***	0.00
	(0.01)	(0.01)	(0.02)	(0.02)
No. of inventors	0.05***	0.06***	0.03	-0.03
	(0.01)	(0.02)	(0.03)	(0.03)
Log($R\&D_{\text{in million \$, in t}}$)	-0.07**	-0.04	-0.13**	-0.09
	(0.04)	(0.04)	(0.06)	(0.07)
Science intensity $_{\text{from t-4 to t-1}}$	-0.02	0.02	-0.08	-0.10
	(0.05)	(0.06)	(0.06)	(0.09)
Basic research focus $_{\text{from t-4 to t-1}}$	1.12	0.20	3.14	2.94
	(0.71)	(0.64)	(1.95)	(2.05)
Size of Pipeline $_t$/$R\&D_{\text{in million \$, t}}$	0.95*	-0.07	2.53***	2.60***
	(0.51)	(0.36)	(0.76)	(0.84)
Size of Pipeline $_t$ /$R\&D_{\text{in million \$, t}}$ ^2	-0.18	0.00	-0.56**	-0.56**
	(0.14)	(0.01)	(0.27)	(0.28)
Size of market for technology	-0.00	-0.00*	0.00	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	11.37*	-56.72***	40.72***	97.45***
	(6.15)	(18.95)	(7.03)	(20.20)
Year dummies	YES	YES	YES	YES
22 Therapeutic field dummies	YES	YES	YES	YES
Biotech compounds dummy	YES	YES	YES	YES
Chemical compounds dummy	YES	YES	YES	YES
Constant	-4.68***	-4.58***	-7.85***	-3.27
	(0.86)	(0.96)	(2.15)	(2.35)
LL	-925.91	-1026.72		-1026.72
N	7836.00	7872.00		7872.00

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A10. Multinomial Logit regression results for take-up limited to patent applications at USPTO

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.31***	0.14	0.50***	0.36**
	(0.09)	(0.13)	(0.13)	(0.18)
D(applied internal SNPRs)	0.62***	0.69***	0.61**	-0.09
	(0.20)	(0.26)	(0.28)	(0.36)
D(basic external SNPRs)	0.20*	0.17	0.20	0.03
	(0.11)	(0.16)	(0.15)	(0.21)
D(applied external SNPRs)	0.53***	0.68***	0.34***	-0.34*
	(0.09)	(0.13)	(0.13)	(0.18)
No. of owners	-0.11	-0.10	-0.05	0.05
	(0.13)	(0.18)	(0.19)	(0.26)
Scope: # IPC6 classes	0.08***	0.08***	0.08***	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.06***	0.06***	0.05***	-0.01
	(0.01)	(0.01)	(0.01)	(0.02)
Log($R\&D_{\text{in million \$, in t}}$)	-0.09***	-0.03	-0.09***	-0.06
	(0.02)	(0.03)	(0.03)	(0.04)
Science intensity $_{\text{from t-4 to t-1}}$	-0.01	0.01	-0.07	-0.08
	(0.03)	(0.05)	(0.04)	(0.07)
Basic research focus $_{\text{from t-4 to t-1}}$	0.46	-0.53	2.79***	3.31***
	(0.38)	(0.38)	(0.76)	(0.84)
Size of Pipeline $_t$/$R\&D_{\text{in million \$, t}}$	0.42***	0.31*	1.79***	1.48***
	(0.13)	(0.18)	(0.32)	(0.36)
Size of Pipeline $_t$ /$R\&D_{\text{in million \$, t}}$^2	-0.03**	-0.01	-0.41***	-0.40***
	(0.01)	(0.01)	(0.12)	(0.12)
Size of market for technology	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	6.52***	-17.86***	16.87***	34.73***
	(2.11)	(6.16)	(2.41)	(6.57)
Year dummies	Yes	Yes	Yes	Yes
22 Therapeutic field dummies	Yes	Yes	Yes	Yes
Biotech compounds dummy	Yes	Yes	Yes	Yes
Chemical compounds dummy	Yes	Yes	Yes	Yes
Constant	-2.83***	-2.84***	-5.74***	-2.89***
	(0.45)	(0.49)	(0.90)	(1.02)
N	13827.00	13827.00		13827.00
LL	-2394.16	-2768.35		-2768.35

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A11. Logit and Multinomial Logit regression results for take-up restricting sample to patents in IPC class A61P

	Model 3: Logit	**Model 4a: Multinomial Logit**		**Model 4b: Mult. Logit**
Reference category		**1.Not taken up**		**2. by other**
		2. by other	**3. by owner**	**3. by owner**
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.30***	0.03	0.53***	0.50**
	(0.11)	(0.16)	(0.14)	(0.21)
D(applied internal SNPRs)	0.72***	0.72**	0.74***	0.03
	(0.20)	(0.29)	(0.27)	(0.38)
D(basic external SNPRs)	0.49***	0.37**	0.59***	0.22
	(0.11)	(0.15)	(0.16)	(0.22)
D(applied external SNPRs)	0.50***	0.58***	0.39***	-0.20
	(0.11)	(0.15)	(0.14)	(0.21)
No. of owners	-0.15	-0.14	-0.09	0.05
	(0.14)	(0.18)	(0.20)	(0.27)
Scope: # IPC6 classes	0.07***	0.08***	0.06***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.06***	0.06***	0.05***	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.09***	-0.05	-0.09**	-0.04
	(0.02)	(0.03)	(0.03)	(0.05)
Science intensity $_{\text{from t-4 to t-1}}$	-0.00	-0.06	-0.01	0.05
	(0.04)	(0.08)	(0.04)	(0.09)
Basic research focus $_{\text{from t-4 to t-1}}$	0.23	-0.49	1.97**	2.46**
	(0.41)	(0.44)	(0.86)	(0.96)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.34**	0.34*	1.45***	1.11**
	(0.16)	(0.18)	(0.44)	(0.48)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	-0.02	-0.02	-0.38*	-0.36*
	(0.02)	(0.01)	(0.21)	(0.21)
Size of market for technology	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	7.36***	-21.66***	18.14***	39.79***
	(2.15	(7.10)	(2.58)	(7.55)
Year dummies	Yes	Yes	Yes	Yes
22 Therapeutic field dummies	Yes	Yes	Yes	Yes
Biotech compounds dummy	Yes	Yes	Yes	Yes
Chemical compounds dummy	Yes	Yes	Yes	Yes
Constant	-3.43***	-3.16***	-6.57***	-3.42***
	(0.47)	(0.51)	(0.96)	(1.08)
N	16144.00	16144.00		16144.00
LL	-2223.92	-2539.75		-2539.75

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A12. Logit and Multinomial Logit regression results for take-up restricting sample to patents in IPC class A61P and (C07 or C12)

	Model 3: Logit	Model 4a: Multinomial Logit		Model 4b: Mult. Logit
Reference category		1.Not taken up		2. by other
		2. by other	3. by owner	3. by owner
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.36***	0.10	0.58***	0.48**
	(0.11)	(0.17)	(0.14)	(0.21)
D(applied internal SNPRs)	0.47*	0.59*	0.43	-0.16
	(0.25)	(0.35)	(0.32)	(0.46)
D(basic external SNPRs)	0.52***	0.41**	0.60***	0.19
	(0.12)	(0.17)	(0.16)	(0.23)
D(applied external SNPRs)	0.48***	0.55***	0.38**	-0.17
	(0.11)	(0.16)	(0.15)	(0.22)
No. of owners	-0.20	-0.22	-0.11	0.11
	(0.15)	(0.21)	(0.21)	(0.29)
Scope: # IPC6 classes	0.07***	0.07***	0.06***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
No. of inventors	0.06***	0.06***	0.05***	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.09***	-0.05	-0.08**	-0.03
	(0.02)	(0.03)	(0.04)	(0.05)
Science intensity $_{\text{from t-4 to t-1}}$	0.00	-0.04	-0.01	0.02
	(0.04)	(0.08)	(0.04)	(0.09)
Basic research focus $_{\text{from t-4 to t-1}}$	0.71	-0.25	2.60***	2.85***
	(0.49)	(0.49)	(0.93)	(1.05)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.26	0.23	1.38***	1.15**
	(0.17)	(0.15)	(0.43)	(0.45)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$^2	-0.01	-0.01	-0.35*	-0.34*
	(0.02)	(0.01)	(0.19)	(0.19)
Size of market for technology	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share in market for technology	8.44***	-19.72***	18.65***	38.36***
	(2.31)	(7.46)	(2.75)	(7.92)
Year dummies	Yes	Yes	Yes	Yes
22 Therapeutic field dummies	Yes	Yes	Yes	Yes
Biotech compounds dummy	Yes	Yes	Yes	Yes
Chemical compounds dummy	Yes	Yes	Yes	Yes
Constant	-3.01***	-3.01***	-5.63***	-2.63**
	(0.59)	(0.66)	(1.10)	(1.28)
N	13385	13385		13385
LL	-1958.01	-1955.89		-1955.89

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Table 1.A13. Logit regression results explaining whether drug candidates are taken up by other firms or by the patent owner

	Model 4	
Dependent variable	**Taken up by other firms versus no take-up**	**Taken up by patent owner versus no take-up**
	Coef. (S.E.)	Coef. (S.E.)
D(basic internal SNPRs)	0.04	0.53***
	(0.16)	(0.14)
D(applied internal SNPRs)	0.64**	0.52
	(0.32)	(0.33)
D(basic external SNPRs)	0.37**	0.55***
	(0.16)	(0.16)
D(applied external SNPRs)	0.59***	0.37**
	(0.16)	(0.15)
No. of owners	-0.11	-0.10
	(0.20)	(0.21)
Scope: # IPC6 classes	0.07***	0.07***
	(0.01)	(0.01)
No. of inventors	0.06***	0.06***
	(0.01)	(0.02)
Log(R&D $_{\text{in million \$, in t}}$)	-0.06*	-0.08**
	(0.03)	(0.04)
Science intensity $_{\text{from t-4 to t-1}}$	0.01	-0.03
	(0.06)	(0.05)
Basic research focus $_{\text{from t-4 to t-1}}$	-0.45	2.46***
	(0.43)	(0.88)
Size of Pipeline $_{t}$/R&D$_{\text{in million \$, t}}$	0.19	1.54***
	(0.15)	(0.39)
Size of Pipeline $_{t}$ /R&D$_{\text{in million \$, t}}$ ^2	-0.01	-0.38**
	(0.01)	(0.16)
Size of market for technology	-0.00	-0.00
	(0.00)	(0.00)
Share in market for technology	-19.28***	18.68***
	(7.17)	(2.68)
Year dummies	YES	YES
22 Therapeutic field dummies	YES	YES
Biotech compounds dummy	YES	YES
Chemical compounds dummy	YES	YES
Constant	-3.43***	-6.58***
	(0.55)	(1.03)
LL	-1194.95	-1275.42
N	18076.00	18109.00

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are reported in brackets.

Chapter 2

Reaching for the Stars

When Does Basic Research Collaboration between Firms and Academic star Scientists Benefit Firm Innovation?[15]

ABSTRACT

While their expertise and scientific excellence make academic star scientists attractive collaboration partners for firms engaging in basic research, the positive consequences of star scientist collaboration for firms' innovation performance are likely to be contingent upon how such collaborations are arranged. Drawing on a dataset of 60 prominent pharmaceutical firms, we only observe an innovation premium of academic star collaboration in basic research (beyond the effect of collaboration with non-stars) if the star scientist abstains from simultaneous collaboration with other firms ('dedication') and is also involved in applied research with the firm ('translation'). If an internal star scientist joins as a collaborator, a translational role of the academic star is no longer a prerequisite and may even be detrimental to an innovation premium.

[15] This chapter is based on a co-authored paper with René Belderbos (KU Leuven, Maastricht University & UNU MERIT), Stijn Kelchtermans (KU Leuven and Hasselt University) and Bart Leten (KU Leuven and Hasselt University)

2.1 INTRODUCTION

It is well-known that the distribution of achievements in scientific research is highly skewed, with a small group of scientists responsible for a disproportional share of research output, in terms of quantity as well as quality (e.g. Lotka, 1926; Rosen, 1981; Azoulay *et al.*, 2010; Kelchtermans and Veugelers, 2011; Azoulay, Stuart and Wang, 2014). These star scientists, as leaders within their research field, also act as important nodes in global scientific networks (Murray 2004; Luo, Koput, and Powell 2009). These characteristics make star scientists interesting hires or collaboration partners for firms in science-based industries involved in intense competition to develop and bring to market innovative products and technologies.

Prior studies have focused attention on the performance benefits that in-house employment of star scientists might bring (P. Almeida, Hohberger, and Parada 2011; A. M. Hess and Rothaermel 2011; A. Hess and Rothaermel 2012; D. Tzabbar and Kehoe 2013; Kehoe and Tzabbar 2015; Grigoriou and Rothaermel 2013; Rothaermel and Hess 2007; Subramanian, Lim, and Soh 2013). However, only a small minority of star scientists is employed by private firms.[16] Star scientists generally appear reluctant to commit themselves to work in the private sector where restrictive company policies may hamper the freedom to publish and define one's own research agenda (Murray 2004; Stern 2004; Aghion, Dewatripont, and Stein 2008). For many firms this leaves collaboration with star scientists employed by universities as the most common strategy to get access to their skills, networks, and expertise.

Extant research has emphasized the importance of basic research in generating innovations in science-based industries (Rosenberg, 1990; Gambardella, 1992; Mansfield, 1995, 1998; Sorenson and Fleming, 2004; Cassiman, Veugelers and Zuniga, 2008; Fabrizio, 2009; Arora, Belenzon and Patacconi, 2015; Della Malva*, et al*., 2015). Prior studies have generally shown positive performance effects of (basic) research collaboration with universities (e.g. Iain M Cockburn and Henderson 1998; Fabrizio 2009; Belderbos, Carree, and Lokshin 2004; Belderbos, Gilsing, and Suzuki 2016b; P. Almeida, Hohberger, and Parada 2011). In the particular context of biotechnology, the research of Zucker and colleagues has indicated a positive role of academic star scientists in the formation and early development of high-tech industries (L G Zucker and Darby

[16] For instance, Zucker et al. (1998a) note that 97.6 % of the star scientists within Genbank were affiliated to a university or research institution.

1996; Lynne G. Zucker, Darby, and Armstrong 1998; Lynne G. Zucker, Darby, and Armstrong 2002b). However, it remains unclear whether established firms can benefit from collaboration in basic research with academic star scientists, in comparison with collaboration with average (non-star) academics. We argue that enjoying these benefits may not be straightforward.

On the one hand, academic star scientists appear particularly attractive partners for collaboration in basic research. Their unique human capital helps deepening firms' fundamental scientific insights, signal firms' scientific excellence and embed firms' research efforts in the larger scientific community. On the other hand, firms may have difficulties in extracting value from collaborating with academic star scientists in basic research. First, star scientists are likely to have a taste for pure and open science, which may conflict with firms' commercialization objectives. Second, given their reputation and academic excellence, and the often-abundant financial resources that come with their status, star scientists have multiple options for industry collaboration and they may utilize this bargaining power to select and shape collaborations for their own interests. When star scientists collaborate with multiple firms, there is an increased risk that knowledge of one collaboration partner spills over to other firms. Third, star scientists' ambitious research agendas may also imply a lack of time and commitment to make substantive contributions to joint research with an industrial partner. Hence, whether collaboration with star scientists in basic research is to be preferred over collaboration with other (non-star) university scientists is likely to depend on how such collaborations are arranged to address potential incongruences with the innovation objectives of the firm.

The current study contributes to the literature on industry-science linkages by examining the conditions under which joint-research with academic star scientists in basic research is associated with an innovation premium in comparison to basic-research collaboration with academic non-star scientists. We argue that three conditions may be of particular relevance. First, the benefits of collaboration may depend on whether the star has a dedicated relationship with the focal firm and abstains from simultaneous collaboration with other firms. Dedication is likely to increase commitment and trust building while limiting risks of knowledge spillovers to potential rival firms. Second, the benefits of collaboration may be greater if the firm can involve the academic star scientist not only in basic research but also in applied collaborative research, aiding the translation of basic research findings into valuable innovations. Finally, as collaborating firms may employ

internal star scientists, involvement of these scientists in collaborative research with academic stars may also affect the success of the collaboration. In particular, we expect that involving internal stars, with their deep knowledge of firms' innovation needs, will benefit the firm if a collaboration with the academic star scientist is lacking a translational (applied research) dimension. If translation becomes the sole responsibility of the internal star scientist, knowledge redundancies and potential conflicts due to ill-defined roles in collaboration with academic stars may be reduced and performance enhanced. Hence, we expect that, within firms' portfolio of basic-research collaboration with academia, partnerships with academic star scientists show a stronger relationship to firms' innovation performance when the academic star scientists show dedication or translation or when internal stars are involved.

Empirically, we analyze a panel dataset (1995-2002) containing detailed information on patents and scientific publications of 60 of the most prominent American, European and Japanese firms in the pharmaceutical industry. We use information on co-publications in basic and applied research journals to measure collaborations between firms and university star scientists and identify star scientists as leaders in their scientific field both in publication and citation performance. We estimate pseudo fixed effects models relating citation weighted patent performance to firms' prior engagement in academic star collaborations under different contingencies, while controlling for a range of relevant firm, star, and collaborative project characteristics. To guide hypothesis development, the quantitative analysis is complemented by a series of interviews with eight academic star scientists in life sciences from two large European research universities and R&D managers from five of the largest European and Japanese pharmaceutical firms involved in firm-university collaboration.

Our results suggest that while basic research collaboration with academia is generally associated with stronger firm-level innovation performance, involvement of academic star scientists shows no outspoken premium on innovation performance within this portfolio of collaborations. Positive effects of academic star scientist collaboration are only observed under specific contingencies. In particular, we find that the benefits from collaboration only materialize for those collaborations where the star scientist abstains from simultaneous collaboration with other firms and where (s) he is involved in both basic and applied research collaboration with the firm. Furthermore, if the basic research collaboration involves an internal star scientist employed by the firm, a translational role

of the external academic star appears not required, and may be even detrimental, for an innovation performance premium. Overall, our findings suggest that, when firms pursue academic collaborations to additionally enhance their innovation performance firms, they should be cautious in their allocation of means across academic collaboration partners and not merely assume an innovation premium from enhanced collaboration with scientists of high excellence.

We note three caveats of our study up front. First, our study focuses on large, vertically integrated pharmaceutical firms, and is not representative for smaller, specialized (biotechnology) firms for which ties to academic star scientists may be of a different nature due to the closeness of the biotech and university communities (Powell, Koput, and Smith-Doerr 1996). Second, although our detailed analyses do not suggest that heterogeneity in project quality, patterns of matching between firms and star scientists, or reverse causality play a role as confounders, the difficulty in finding suitable instruments for our focal variables precludes us from clearly separating selection from treatment effects and interpreting our findings as causal relationships. Our analyses, for example, do not fully cover the matching process between firms and stars. Omitted firm and star characteristics influencing the matching process and related to the variables of interest may pose alternative explanations for our findings. However, the findings are robust to inclusion of the main matching characteristics identified in prior literature as shown in the supplementary analyses (Mindruta, 2013; Banal-Estanol, Macho-Stadler and Perez-Castril, 2014). Third, the study focuses on characteristics measured at the level of individual academic star scientists while team effects are not taken into account. As it is common practice within life sciences that top researchers obtain their own lab and that lab heads are included as last author on each publication of their lab (Lissoni, Montobbio, and Zirulia 2013; Lissoni and Montobbio 2015), we do believe that the research output of academic stars approximates the expertise of and firms' collaborative relationship to the star team. Despite these caveats, our results show robustness to multiple supplementary analyses, like the inclusion of prior collaborative experience with the academic stars. To conclude, we strongly believe that this study provides important insights to the literature emphasizing the difficulties of collaboration with academic stars and to practitioners noting the importance of careful selection and management of academic collaborations.

2.2 THEORETICAL BACKGROUND AND HYPOTHESES

2.2.1 Basic research and collaborations with academia

Basic research is an important driver of innovation in science-based industries (Mansfield 1995; Narin, Hamilton, and Olivastro 1997). Innovation-seeking firms systematically search for fundamental insights to conduct well-informed experiments and to identify promising directions for new research (Rosenberg 1990; Cassiman, Veugelers, and Zuniga 2008; Fabrizio 2009; Sorenson and Fleming 2004). Numerous empirical studies have shown the importance of basic research in improving firms' innovation performance (e.g. Rosenberg, 1990; Gambardella, 1992; Cockburn and Henderson, 1998; Sorenson and Fleming, 2004).[17]

Prior research has also suggested that the benefits of performing basic research are greater when it is conducted in collaboration with universities (Cockburn and Henderson, 1998; Lynne G. Zucker, Darby and Armstrong, 2002; Fabrizio, 2009). As firms' R&D-departments find it difficult to remain up-to-date with all relevant scientific advances, firms turn to university partners to provide guidance and scientific expertise in research areas relevant to the firm (Cassiman and Veugelers 2006).

At the same time, however, it is clear that engagement in university-industry collaboration also faces numerous obstacles that need to be overcome. The defining characteristics of academic science, such as the rapid disclosure and wide dissemination of research results as well as the recognition-based reward system, are different from the performance goals and incentive systems in firms (Arora and Gambardella 1994). The contrasting views that academia and industry have on science can lead to conflicting research goals and priorities for joint research projects (Bruneel, D'Este, and Salter 2010; V. Tartari, Salter, and D'Este 2012). Collaborations in basic research in this regard pose particular difficulties, as they require close interaction and understanding to transfer tacit and complex knowledge across organizations (Bruneel, D'Este, and Salter 2010; Plewa et al. 2013; V. Tartari, Salter, and D'Este 2012). The high levels of uncertainty that are defining basic research, combined with difficult-to-monitor knowledge generation and transfer, rule out complete contracting to govern basic research collaborations. Instead, trust and mutual interdependence are crucial (Faems, Janssens, and Van Looy 2010). Another factor complicating the governance of firm-university research collaborations is the bargaining over IP ownership

[17] The study by Lim (2004) is an exception and does not identify positive performance benefits.

(Belderbos et al. 2014). IP negotiations are non-trivial exercises since the knowledge contributions of the respective parties are sometimes difficult to pin down and because universities increasingly pursue commercial aims leading to difficult negotiations on the ownership of jointly developed IP (Bruneel, D'Este, and Salter 2010).

2.2.2 The value and challenges of collaborative basic research with academic star scientists

Among university-industry collaborations, academic star scientists may be particularly attractive research partners because of the extraordinary human and social capital they bring to such partnerships. For instance, firms' timing and location of entry in the emerging area of biotechnology was strongly related to their collaboration with academic star scientists (Lynne G. Zucker, Darby, and Armstrong 1998). Moreover, the interviewed R&D managers were unanimous on the principal selection criteria of 'excellence':

> *"(When selecting an academic research partner), I think the priority would be on the individual and their scientific excellence."*
> *"We try to put the science first. We want to be convinced that what we are looking at is really compelling, state of the art and transformational"*
> *"What matters is scientific excellence. That is really the number one criteria (for selecting a research partner)"*

Beside excellence, the interviewed managers and scientists referred to informal contact and past collaborative experience as enablers of collaboration. Also, past research has shown that peers with industry contacts stimulate scientists to step into university-industry collaborations themselves (Perkmann *et al*., 2013a). Moreover, past collaborative experience has been mentioned in multiple studies as a facilitator of collaboration (e.g. Bruneel, D'Este and Salter, 2010; Tartari, Salter and D'Este, 2012). Empirical studies on the matching between academic scientists and firms have presented higher collaboration likelihoods between partners of similar research quality (Banal-estañol, Macho-stadler, and Pérez-castrillo 2017; Mindruta 2013), similar focus on basic research (Banal-estañol, Macho-stadler, and Pérez-castrillo 2017). Moreover, they signaled that specialized partners are matching with partners with a broader research diversity (Mindruta 2013). While controlling for the research characteristics of both the firm and academic star scientists, this study will focus on the importance of 'excellence' among academic collaboration partners for firms who aim to increase their innovation performance through these collaborations. More precise, we will

analyze the relationship between firms' patenting performance and the extent of collaboration with academic star scientists within firms' portfolio of basic-research collaborations with academia.

Firms' innovation performance may benefits from collaboration with star scientists in several ways. First, stars may convey valuable tacit knowledge beyond what is codified in journal articles (Arora and Gambardella 1990; Iain M Cockburn and Henderson 1998) and they may disclose preliminary research results, on which collaborating firms can build their own applied research, faster than rival firms without access to the star (Fabrizio 2009). Second, given the expertise and deep knowledge of the academic star scientist, collaboration may be instrumental to enhance the quality of firms' basic research and their understanding of the relevant technological landscape (Gambardella 1992), helping them in the selection of fruitful research avenues, thus avoiding costly research trials (Cassiman, Veugelers, and Zuniga 2008; Fabrizio 2009; Sorenson and Fleming 2004; Rosenberg 1990). Third, interaction with academic star scientists can enhance the research capabilities of firms' R&D departments, not only by helping them to interpret results of internal research (Rosenberg 1990) but also to identify and understand the results and implications of externally conducted basic research (Cockburn and Henderson 1998; Gambardella 1992). Fourth, academic star scientists occupy central positions in international research networks and have large networks of research partners (A. Hess and Rothaermel 2012) which can be activated in collaborative projects with firms. Based on these arguments, it can be expected that within firms' portfolio of collaborations with academia, partnerships with academic stars are most strongly associated with firms' innovation performance.

However, collaboration with star scientists also poses significant challenges. These include the challenges common to all firm–university collaborations: overcoming the differences in work practices and incentives between universities and industry, achieving trust between partners, reaching IP agreements and dealing with potential knowledge spillovers to other firms given the public good characteristics of knowledge. Firms' collaborations with star scientists are likely to face additional difficulties. A first relevant characteristic of such collaborations is the stronger independence of academic star scientists relative to non-stars. Star scientists tend to have good access to support from funding agencies and collaboration partners from academia and industry, and are less likely to depend on a single collaboration partner for funding or the realization of their research projects. Hence, they can be more selective in choosing their research partners (Audretsch

and Stephan 1996) and will be better able to negotiate collaboration contracts that meet their own interests. One of the interviewed star scientists illustrates this:

> *"They [firms] can collaborate with us if they are interested, and if they are not, then we will just proceed with other partners, if necessary."*

One of the interviewed R&D managers confirmed the difficulty of attracting star scientists for industrial collaboration:

> *"Big names can be quite demanding and have got high expectations. […] Sometimes it's much more pleasant to work with young and upcoming professors who haven't made their name yet, but are very open and eager to work with others".*

Second, star scientists often take up additional managerial responsibilities and tasks in addition to their broad research portfolio, limiting available time to perform collaborative research with firms. In the life sciences, star scientists often manage large research laboratories, involving responsibilities such as Ph.D. supervision, funding acquisition, and people management. Furthermore, star scientists are often involved in editorial work for journals and frequently travel to and attend international events and conferences. The time pressure resulting from taking care of these broader responsibilities may force star scientists to limit their efforts and commitment to collaborative research project, which may harm the contribution of the project to the innovation performance of the collaborating firm.

Below we explain that these hurdles characterizing firms' collaboration in basic research with academic star scientists are likely to imply important contingencies for firms to fully benefit from these collaborative efforts within their portfolio of basic-research collaborations with academia. More specifically, we will argue that the innovation benefits of firms' collaborative efforts with academic stars can increase by engaging in 'dedicated' collaborations – where the star scientist abstains from simultaneous collaboration with other firms – and 'translational' collaborations – where the research collaboration is not limited to basic research but extends to applied research. With translational research, we refer to a process in which basic research, through follow-up applied research, produces new drugs and devices for treating patients (Woolf 2008). Hence, we do not use the broader perspective as defined by the European Society for Translational Medicine

in which 'translational' refers to interdisciplinary research, which also includes, for example, the optimization of systems of care (Cohrs *et al.* 2015). Finally, we argue that the co-involvement of internal star scientists in collaborative research with academic star scientists only benefits the firm if the academic star scientist cannot fulfill the translational role.

2.2.3 Dedicated collaboration with academic stars

Within firms' portfolio of academic collaborations, dedicated collaboration with academic stars, where the focal firm is the only industrial research partner of the academic star scientist, may not only represent innovation benefits of excellent human capital but may also alleviate concerns and difficulties pertaining to collaborative research with academic star scientists. These arguments relate to the threat of knowledge spillovers, the time constraints star scientists face, and trust building.

First, an important issue a firm has to deal with when working with academic scientists, is the partial public good nature of scientific knowledge (Arrow 1962; Nelson 1959). While the development of knowledge requires significant investments, knowledge spillovers can lead to competitors' free riding on these investments at limited learning costs. Moreover, these learning costs drop considerably with proximity to the scientist possessing such knowledge (Zucker *et al.,* 2002). Hence, rival firms engaged in parallel research projects with the same academic star scientist may experience significantly lower learning costs, and may pose a serious threat to the collaborating firm in the race to establish patented inventions. Even contractual limitations on sharing certain pieces of knowledge developed in collaboration with a star may not completely avoid knowledge from spilling over, since the transfer of knowledge - in particular if it is tacit in nature - is hard to monitor and, consequently, contract breaches are hard to legally enforce. The interviewed R&D managers recognized the risk of knowledge spillovers:

"It's sometimes a very thin line to know which information scientists really need to have and how far do you go in sharing information. It's finding the right balance to create trust and have a very collaborative environment in which you can both operate and exchange information, but not going beyond what is essential for both parties to do what they're supposed to be doing and not turn them into a competitor."

> *"You have to assess very carefully what that person's level of involvement with a competitor is."*

If the university star scientist solely collaborates with the focal firm, this may considerably limit the probability of knowledge spilling over to competing firms, as there is no parallel knowledge exchange with other firms. Even if other firms face only temporarily restricted access to the academic star scientist this may help the focal firm to gain a competitive advantage in patent and drug development races. Without direct interactions with the academic star scientist, other firms are less likely to obtain the tacit knowledge to put the results of basic research to productive use (Cohen and Levinthal 1989; Arora and Gambardella 1990). Similarly, the dedicated star scientist will be less restricted in her communication with the scientists of the focal collaborating firm as she is not hindered by secrecy agreements stemming from other projects, and there is less need to worry about inadvertently disclosing confidential information across projects.

A second hurdle to successful firm-academic star collaboration is the independence of academic star scientists and the alternative opportunities they have to pursue research and to get funding. Star scientists can make strong demands during negotiations. They may negotiate collaborative contracts that put an upper bound on the time they invest in the collaboration, or they may economize on time invested once the collaboration is in place. An interviewed star scientist expressed this tendency to economize on time:

> *"Well, my preference would be the least amount of face-time. I mean, frequent meetings chew up a vast amount of my week and having to attend an extra meeting or a regular meeting would be a negative."*

The competition for a star's time was also raised by an interviewed R&D manager:

> *"If they're very busy and collaborating with a lot of others there will be time constraints".*

Time constraints are likely to make it more difficult for a focal firm to engage in frequent and profound interactions with the star scientist that are conducive to successful collaboration. While collaborations among academic partners tend to be more flexible and informal, collaboration with industrial partners is often subject to strict time management with contractually determined deadlines and milestones. Hence, if the star scientist has multiple industrial collaboration partners

competing for her attention, it will be more difficult to devote sufficient time to each collaboration partner, which may reduce effective learning effects and collaboration benefits for the firms.

Finally, dedicated collaborative relationships are more likely to be characterized by improved knowledge sharing and trust. Firms that enjoy dedicated access to a star scientist are likely to commit more strongly to the relationship (Colyvas et al. 2002) and invest in relationship-specific assets (e.g. Elfenbein and Lerner, 2012) due to the privileged access and associated better knowledge appropriation prospects. The increased availability and commitment of both the star scientist and the firm, and the reduced concerns about unwanted knowledge disclosures, may lead to stronger interpersonal relationships and the buildup of trust, mutual understanding, and goodwill (Granovetter, 1985; Coleman, 1988). Strong social relationships in turn enhance the depth of knowledge sharing and the effectiveness of research collaborations (Bruneel *et al.* 2010; Plewa *et al.* 2013; Tartari *et al.* 2012).

The above arguments suggest that dedication may mitigate problems associated with academic star collaboration and strengthen potential benefits. Hence, considering the project and relationship specific costs for each collaboration, firms' innovation performance may be higher when more of its star-firm collaborations are accomplished with dedicated academic stars. This leads to the following hypothesis:

Hypothesis 1 (Dedication): *The innovation premium a firm may reap from basic research collaborations with academic star scientists is positively associated with the degree to which these collaborations involve dedicated stars (star scientists with no other industrial collaboration partner than the focal firm).*

2.2.4 Translation: extending basic research collaboration to applied research

While collaboration with dedicated academic stars may improve knowledge sharing between the academic star scientist and the firm, the involvement of the star in collaborative applied research – besides basic research – may also be a vital element in successfully collaborating with academic star scientists.

A key challenge for firms is the 'translation' of basic research to applied research focusing on successful technology development with commercial orientation. Applied research benefits from

tacit knowledge related to the basic research process, such as information on the trial and error process that has led to the scientific findings, which will be relevant information to establish critical conditions for successful experiments (Agrawal 2006; Fuller and Rothaermel 2012; Sorenson, Rivkin, and Fleming 2006). Knowledge on the basic research process resides in the minds of the discovering scientists who have an information advantage towards others in implementing this knowledge into successful innovations (Jensen, Thursby, and Thursby 2003). As the translation process is risky and highly uncertain, it is impossible to determine in advance when and what knowledge a firm may need during the applied research process. It will therefore be important for firms to stay in close contact with the discovering academic star scientists to discuss solutions if applied research faces obstacles. Hence, involving the star scientist also in applied research is likely to have important benefits for the collaborating firm. One of the interviewed star scientists referred to his deep understanding of the invention process as a reason to remain involved during applied research (drug development):

> *"[we may remain involved in drug development because] we are sometimes the experts who are more knowledgeable on how something works, how the drug should be developed. From their side, they [the corporate scientists] are of course the experts in technically realizing the development. In the best case, it [the drug development] is accomplished in collaboration."*

The interviewed R&D managers also referred to the benefits of involving academic star scientists in both basic and applied research:

> *"The academic partner may have some very deep, fundamental knowledge about something and we understand the whole process of discovery and development better. There is a mutual enrichment at work here. I would say in the long run it's good to keep the originating principal investigator on board as long as possible. It might be that certain questions in pre-clinical development actually can be addressed with the basic research know-how that a person has."*

In general, firms face the problem that excellence in basic science does not automatically imply valuable innovation. The criteria of what constitutes good academic research and what defines a

successful innovation are not the same (Foray and Lissoni 2010; Gittelman and Kogut 2003). Academic star scientists who engage in both basic and applied research with a focal firm get acquainted with both perspectives and develop competences to recognize opportunities in basic research that are worth to pursue further by the focal firm. Although star scientists differ in their 'taste' for applied research and commercialization (Sauermann and Roach 2012; Sauermann and Stephan 2013; Stokes 1997) some stars are willing to engage in collaborative applied research, as this may provide valuable feedback for their basic research and allows them to see their research materialize in actual therapeutic or diagnostic results (Martin Meyer 2006; Rosenberg 1990; Sauermann and Roach 2011). The latter motivation is illustrated by a quote from a star scientist[18]:

> *"Working with [biotech firm] Oncurious gives me the occasion to remain very closely involved in the process of developing a drug for clinical use."*

By collaborating with the focal firm in applied research, the star scientist does not only strengthen her understanding of research translation, but also learns what the development expertise and applied research skills of the focal firm are. Through these insights, translational star scientists will better recognize specific research and development opportunities in collaboration with the firm and will therefore be better able to tailor the collaboration to the firm's needs.

The above arguments and insights suggest that an academic star scientist who can span the basic-applied boundary within the setting of collaborative research can align her expertise in basic research with the development practices and innovation needs of the firm. Hence, firms, who aim for the largest innovation benefit from their basic-research collaborations with academic stars, are expected to establish a larger share of these collaboration with translational academic stars.

Hypothesis 2 (Translation): *The innovation performance premium a firm may reap from basic research collaborations with academic star scientists is positively associated with the degree to which these collaborations are translational (the star scientists are also involved in applied research collaboration with the focal firm).*

[18] Quote by Prof. Dr. Carmeliet, winner of the Ernst Jung Prize in medicine (Oncurious 2015).

2.2.5 Involving internal star scientists in academic star collaboration

While we consider firms' collaboration with academic star scientists, extant work has focused on internal star scientists. Like their academic counterparts, these internal stars embody unique human capital and they excel in terms of their scientific contribution. Prior studies have reported both positive (P. Almeida, Hohberger, and Parada 2011; Subramanian, Lim, and Soh 2013), negative (Zucker *et al.* 2002) and insignificant (Rothaermel and Hess 2007; A. M. Hess and Rothaermel 2011; A. Hess and Rothaermel 2012) effects of internally employed star scientists on firms' innovation performance. We complement this prior work on the general performance implications of internal star scientists by examining the effects of the involvement of internal star scientists in collaborative basic research with academic star scientists.

There are both positive and negative influences to be expected of involving internal star scientists in basic research collaboration with external academic star scientists. On the one hand, having a star scientist on both sides of the partnership may increase mutual understanding and trust. The balance in terms of scientific excellence may improve the understanding and absorption of the knowledge and research expertise that either side brings to the table, leading to more effective communication and collaboration. This is especially the case when knowledge is characterized by a high level of tacitness and knowledge transfer requires close personal interactions and working together on the same research project (Zucker *et al.*, 2002). The solid scientific reputation of both the internal and external star scientists is also likely to increase mutual trust and respect, which stimulates knowledge sharing among the star scientists and improves the effectiveness of research collaborations (Bruneel et *al.* 2010; Plewa *et al.* 2013; Tartari *et al*. 2012). The interviewed scientists and R&D managers experienced these communication benefits:

> *"There are a couple of key advantages [of working with an internal star scientist]. One is they speak the same language. […] [Second,] they also understand more of your [the academic star's] motivations and what it is that you want out of a collaboration."*

> *"[Collaborating with a star scientist within the company] turns a one way interaction into a two way interaction."*

> *"They [the academic star scientist and the internal scientists] need to understand each other well. That is why we [the company] have an expert talking to a university expert."*

On the other hand, involving internal star scientists in collaborative research with academic star scientists may also result in knowledge redundancy or, in the worst case, conflict. First, academic star scientists possess the research expertise and access to the academic community that internal star scientists can also bring to a research collaboration. Hence, both types of star scientists can be considered as substitutive sources of knowledge in a collaboration (Hess and Rothaermel 2011; Subramanian *et al.* 2013) and the marginal benefit of an academic star scientist may be smaller if an internal star is also involved. Second, accommodating top performers within the same team may lead to inefficiency and even conflict (e.g. Groysberg *et al.*, 2011; Cattani *et al.*, 2013). Teams tend to benefit from some hierarchy and clear roles as this brings clarity to social interaction, assigns accountability for task accomplishment and sets rules for the distribution of resources within the group. In teams with more than one star, a clear hierarchy is missing and egos may get in the way of decision making and knowledge sharing (Groysberg *et al.*, 2011). During the interviews, the importance of complementarity in skills and knowledge came forward:

> *"If we [academic star scientists] look for a collaboration partner within a company then I would more likely search for someone who can make the difference within the company instead of someone with a similar profile.*

> *"If we [the academic star scientist and the internal scientist] would have the same knowledge and skills, there would be no need for a collaboration. […] It is the lack of knowledge and skills that forms the basis of collaboration, otherwise you can do it yourself."*

The arguments above do not suggest an unambiguous effect of internal star scientist involvement in collaborative research with academic star scientists. While there appears no reason to expect that the consequences of the co-involvement of internal stars relate to the dedication of the external star (potential conflict or knowledge redundancy remain), we do posit that the presence or absence of a translational element (applied research) in the collaboration with the academic star scientist may matter. An internal star scientist is well positioned to conduct translational research, in particular when she was involved in the underlying (collaborative) basic research. An internally employed

scientist is acquainted with the firm's needs and research practices and typically has applied research affinity and experience. Hence, in the absence of an applied research element in the collaboration with the academic star scientist, the internal star is well positioned to assume this translational role instead. This may involve a specialization of tasks amongst both types of star scientists, with the internal star taking on the translational role, participating in the basic research collaboration from that perspective. With such task specialization, knowledge redundancy in the collaboration and conflicts due to ill-defined roles are less likely to occur, while the involvement of internal and external star scientists may provide benefits in terms of deeper scientific understanding and mutual trust.

Hypothesis 3 (academic and internal stars): *The innovation performance premium a firm may reap from basic research collaborations with academic star scientists that also involve internal star scientists of the firm, is positively associated with the degree to which the academic star scientist collaborations are non-translational (the academic star scientists are not involved in applied research collaboration with the focal firms).*

2.3 DATA, VARIABLES AND METHODS

We test our hypotheses by relating the innovation performance of pharmaceutical firms to their past (collaborative) basic research activities and the characteristics of such collaborations. Collaboration may stimulate innovation performance not only directly through the development of project-specific patents based on the collaborative research, but also through more general knowledge transfer and learning, affecting the broader R&D portfolio of the firm (e.g. Cassiman *et al.* 2008). In other words, while collaboration with star scientists may serve multiple short- and long-term goals of the firm, we focus on the effect on firm-level innovative performance.

Firms select academic stars to collaborate with, and stars select firms. The selection process underlying star-firm collaboration of various types may lead to different innovation outcomes due to the specific characteristics of the academic star, firm, and project, which may bias inferences on the role of the hypothesized contingencies. We address this by estimating elaborate models controlling for a range of firm characteristics, star characteristics, and characteristics of the collaborative research projects which are likely to be relevant in the selection process and may affect innovation performance. In particular, the analysis controls for prior star-firm collaboration,

which may influence selection and may at the same time improve the effectiveness of collaboration, and an indicator of the quality of the collaborative projects, which could be associated with the choice for specific project types. In a supplementary analysis, we also examine the robustness of our findings when controlling for particular characteristics of a firm and its academic partner that may make matches more productive (Mindruta 2013; Banal-estañol, Macho-stadler, and Pérez-castrillo 2017).

2.3.1 Data

We constructed a panel dataset on the patent and publication activities of 60 of the most prominent pharmaceutical firms in the world from 1995 to 2002. The firms have headquarters in the United States, Europe or Japan and are the largest R&D spenders (in absolute terms) in the pharmaceutical industry as reported in the 2004 EU Industrial R&D Investment Scoreboard. This ranking lists the top 500 corporate R&D investors based in Europe, and the top 500 companies based outside Europe (mainly in the US and Japan), in 2003. A list of the sample firms is provided in the appendix. We rely on an unbalanced panel dataset of 406 observations in our empirical analyses, as some sample firms were only created after 1995 (e.g. Novartis was formed as a result of a merger of Ciba-Geigy and Sandoz in 1996) and due to some missing values for R&D expenses.

The characteristics of research by the major players in the pharmaceutical industry makes this a particularly interesting context for investigating academic star-firm collaborations. The science-based nature of research in the pharmaceutical industry and the high patenting and publication rates allow examining research collaboration processes and outcomes through quantitative analysis. The major pharmaceutical companies are involved in the entire process from basic research to drug development and clinical trials (Campbell 2005), with the interplay between basic and applied research being an important aspect of firm performance. Finally, the ongoing debate on the productivity crisis in the pharmaceutical sector (Rafols et al. 2014) calls for deeper insights on the characteristics of effective research (collaboration) strategies.

Innovation performance: patent data.

We utilize patent data to measure firms' innovation performance. Patent data is extracted from the PATSTAT database (2011 update), which contains information on patents from all major patent offices worldwide. Patents are a good indicator in our context as the propensity to patent inventions is high in the pharmaceutical industry (Arundel and Kabla, 1998). We weight patent counts by the

number of forward patent citations to control for differences in the economic value of patents (Trajtenberg 1990; Gambardella, Harhoff, and Verspagen 2008; Hall, Jaffe, and Trajtenberg 2005). Recent research (J.-Y. Chiou et al. 2016; J. Chiou et al. 2012) on molecular entities indicates a strong correlation between the citation rate of pharmaceutical patents and the successful introduction of drugs based on these patents. The citation-weighted patent count is based on a fixed 4-year window of forward citations to establish a comparable citation window across patents (Hall, Jaffe, and Trajtenberg 2005; Trajtenberg 1990). For the calculation of citation counts, both the citing and cited patents are integrated at the DOCDB PATSTAT patent family level to avoid double counting patents on similar innovations (Martinez 2011).

Patent data are collected at the consolidated parent firm level by searching for patents under the name of the parent firm as well as all their majority-owned subsidiaries. For this purpose, yearly lists of companies' subsidiaries were used, as reported in corporate annual reports, yearly 10-K reports filed with the SEC in the US, and, for Japanese firms, information on foreign subsidiaries published by Toyo Keizai in the yearly 'Directories of Japanese Overseas Investments'. Acquired firms and their patent stocks are considered part of a parent firm from the year the acquisition transaction was completed.

Basic research: publication data.
We draw on information contained in publication data on pharmaceutical research in the PubMed database to construct indicators of firms' research activity and collaborative behavior. Publication counts are strong and timely indicators of firms' levels of involvement in research in science-based industries (Arora and Gambardella 1990; Gambardella 1992), since the turn-around time of publications in life sciences is typically only a few months (Kaplan, Murray, and Henderson 2003; Murray and Stern 2007). We relate innovation (patent) performance of firms to (collaborative) basic research activity in the past 4 years (t-4 to t-1). This time window is likely to be sufficient, as most patents are applied for relatively early in the drug discovery phase (Campbell 2005). Since we are interested in firms' (collaborative) basic research focusing on compound (drug) discovery we use the CHI journal classification scheme to distinguish applied research (levels 1 and 2) from basic research (levels 3 and 4) (Hamilton 2003; Thursby and Thursby 2011).

Collaboration with (star) academics: co-publications.
We draw on information from co-publications to build collaboration measures for the sample firms (Cockburn and Henderson 1998; Fabrizio 2009). We identify academic co-authorship by a string-

matching algorithm that recognizes affiliations of universities or research institutions. Prior research has validated co-publications as a reliable indicator of collaborative research (Cockburn and Henderson 1998; Fabrizio 2009). Most collaborations result in co-authored publications (Melin and Persson 1996), and most co-publications do reflect actual research collaborations (Hicks, Isard, and Martin 1996).

Among the academic co-publications of firms, we identify collaborations with academic star scientists. We draw on disambiguated author names in the Authority dataset of Torvik and colleagues (Torvik and Smalheiser 2009; Torvik et al. 2005), which has uniquely identified authors on PubMed publications. The authors of the firm (co-)publications can be compared with all authors within PubMed on the basis of their complete publication record. We follow the definition of Rothaermel and Hess (2007, 2012) and identify star scientists as those authors whose publication and citation performance are both three standard deviations above the means in their scientific field. We apply the criterion in a dynamic manner, using a moving 4-year window to allow for changes in star scientist status due to retirement or career changes (Boris Groysberg, Nanda, and Nohria 2004; Boris Groysberg, Lee, and Nanda 2008). Star scientist status is assessed per scientific field to control for discipline-specific publication and citation patterns (Stijn Kelchtermans and Veugelers 2013). Based on the journal categorization of Thomson Reuters in 2014, we consider 44 distinct fields in the scientific domains "Medicine" and "Life Sciences". If the scientist has, in at least one of these fields, publication and citation counts that both exceed the aforementioned threshold she is considered a star scientist.

In total, from among 2,478,517 scientists with at least one publication in PubMed in the broader fields of "Medicine" or "Life Sciences", we identified 26,586 (1.1%) star scientists. For the sample firms, among the 126,325 authors listed on their basic research publications, 7,340 (5.8%) were identified as stars. This set of star scientists consists of both scientists employed internally by the firms and those employed by a research institution or university. To identify star scientists working in academia we applied specific string-matching algorithms on three types of affiliation data: first-author addresses, email-addresses (Torvik *et al.* 2005; Torvik and Smalheiser 2009) and addresses listed on the corresponding Web of Science publications of the PubMed publications. In line with previous studies (e.g., Zucker *et al.* 1998), we find that most star scientists listed on the publications of the sample firms are academics (6,554 or 89%). Among the 60 firms in the sample, 34 firms

also employ internal star scientists, who author a total of 7,681 basic research publications. Of these, 2,172 are in collaboration with an academic star scientist. The sample firms' basic research collaborations with academic star scientists not involving internal stars is about six times that number (13,622).

An academic star-firm basic research collaboration is considered "translational" if the collaboration not only includes joint basic research but if the star scientist is also involved in applied research with the focal firm in the same 4-year window. An academic star-firm collaboration is categorized as "dedicated" if the star scientist is not mentioned as co-author on any publication (basic or applied) of another (bio) pharmaceutical firm during the observation period of 4 years. We consider the 60 largest R&D spending pharmaceutical firms in the sample in addition to 76 of the largest R&D spending biotechnology firms in 2004 as the relevant players within the (bio) pharmaceutical industry to determine whether a star is "dedicated" to a single focal firm.

As dedication and translation are not mutually exclusive characteristics of collaboration, we identify four types of collaboration with academic stars: dedicated and translational, dedicated and non-translational, non-dedicated and translational, and non-dedicated and non-translational. The set-up with four exclusive categories provides the most detailed insights into the performance effects of different configurations of dedicated and translational collaborative research, allowing to disentangle their individual influences. We furthermore distinguish combinations of these four collaboration types with or without the involvement of internal stars, such that we arrive at eight exclusive categories of academic star collaborations in our full regression model.

The construction and descriptive statistics of the collaboration variables are shown in Figure 1. About 58% of firms' basic research publications are co-authored with academics. On average, a quarter of these collaborative basic research publications are co-authored by an academic star, while about three percent of these publications involve only internal stars. Non-star collaborations form the majority of basic research collaborations with academia (72%). Among the collaborations involving academic stars, about 8 percent also involve internal stars. If we further disentangle these two categories of academic star collaborations, we see that in case of involvement of academic stars only, the collaboration type 'non-dedicated and non-translational' is the most common (60.1%), followed by dedicated and non-translational (23.8%), non-dedicated and translational (10.9%), and dedicated and translational (5%). For the academic star collaborations also involving

internal stars, the shares of the four types of collaborations are broadly comparable, although the share of dedicated translational collaborations is higher at roughly 10%. Figure 2 presents the key variables across different levels of observed innovation performance visualizing the topic of interest of this study. Also here it is noticeable that the non-translational and non-dedicated collaboration are most common at all levels of performance. Collaborations with only academic stars who are translational and dedicated appear to be least frequent for firms performing just above the median. The other forms of firm-star scientist collaborations present a more stable and steady growth.

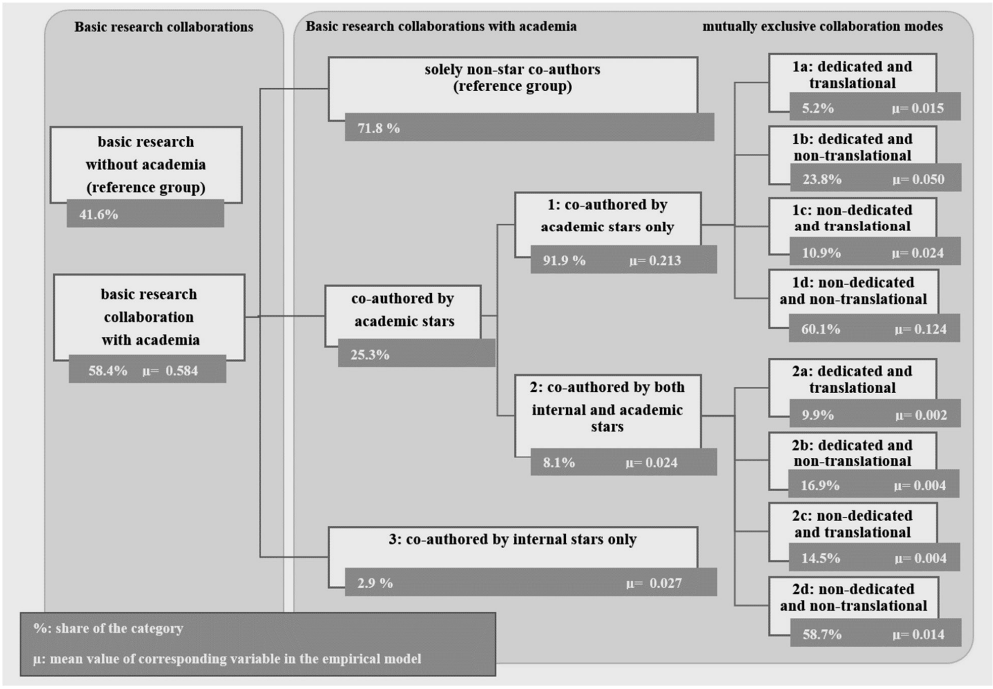


Figure 2.1. Construction and descriptive statistics of the key collaboration variables

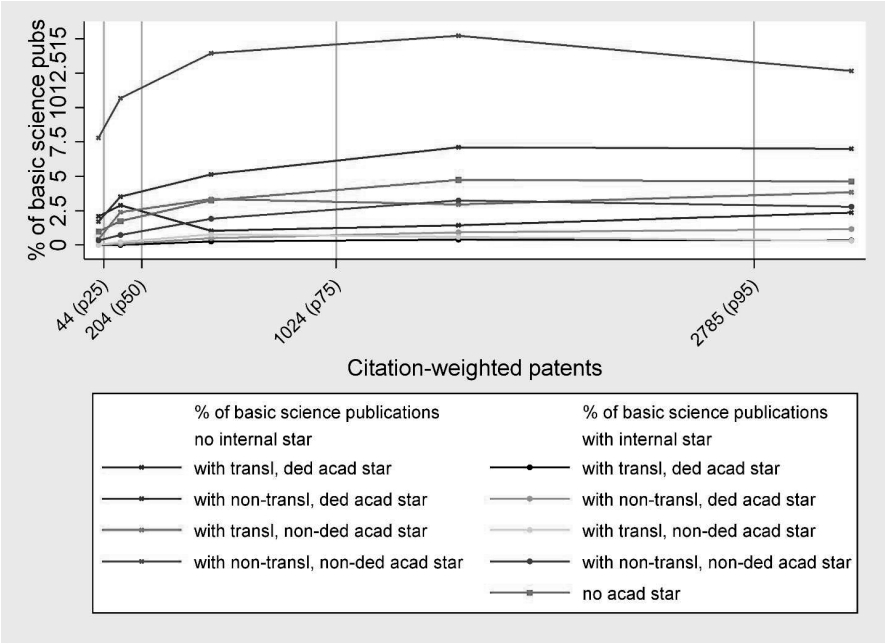


Figure 2.2. the key variables of basic research collaboration across observed innovation performance

2.3.2 Empirical model and variables

We opt for count data models as they take into account the non-negativity and discreteness of our dependent variable: firms' citation weighted patent count. We estimate quasi maximum likelihood Poisson models that are robust to over-dispersion and against distributional misspecification (Cameron and Trivedi 2009). To control for time-invariant heterogeneity across firms that is not captured by the model variables, we estimate pseudo-fixed effects models by including the pre-sample average of the dependent variable. The advantage of pseudo fixed models is that they do not require strict exogeneity of the error terms, as is the case for conventional fixed effects models (Blundell, Griffith, and Van Reenen 1999; Blundell, Griffith, and Vanreenen 1995). The pseudo fixed effect is measured as the natural logarithm of the firms' patents 5 to 8 years before the observation period. This relatively long pre-sample period avoids convolution with the explanatory variables in the model, such as the basic research collaborations measured 1 through 4 years prior to the first observation on innovation performance.

Table 2.1. Descriptive statistics (406 obs.)

	Name	Mean	S.D.	Min.	Max.	# firms with value >0
DEPENDENT VARIABLE						
1	Citation weighted patent count $_{t}$	644.628	893.910	0	4,551	60
KEY VARIABLES						
2	Log (R&D $_{t-1,\ in\ million\ USD}$)	5.216	1.692	0.196	8.486	60
3	Basic research publications$_{t-4\ to\ t-1}$/ R&D$_{t-1,\ in\ million\ USD}$	0.771	0.977	0.000	11.890	59
4	Basic research co-publ. with academia $_{t-4\ to\ t-1}$ /Basic research publications $_{t-4\ to\ t-1}$	0.548	0.214	0.000	1.000	57
	Share of basic research co-publications with academia $_{t-4\ to\ t-1}$					
5	… with academic stars	0.236	0.155	0.000	1.000	55
	… without internal stars and with academic stars	0.213	0.138	0.000	1.000	55
6	… which are dedicated and translational	0.015	0.062	0.000	1.000	39
7	… which are dedicated and non-translational	0.050	0.051	0.000	0.333	47
8	… which are non-dedicated and translational	0.024	0.036	0.000	0.245	43
9	… which are non-dedicated and non-translational	0.124	0.094	0.000	0.571	53
	…with internal stars and with academic stars	0.024	0.047	0.000	0.250	29
10	… which are dedicated and translational	0.002	0.006	0.000	0.045	13
11	… which are dedicated and non-translational	0.004	0.010	0.000	0.092	20
12	… which are non-dedicated and translational	0.004	0.014	0.000	0.139	20
13	… which are non-dedicated and non-translational	0.014	0.032	0.000	0.162	27
14	…with internal stars and without academic stars	0.027	0.074	0.000	0.250	32
FIRM CONTROLS						
15	Technological diversity $_{t-4\ to\ t-1,\ IPC3}$	6.867	3.044	0.000	19.151	60
16	Research alliances $_{t-4\ to\ t-1}$ / R&D $_{t-1,\ in\ million\ USD}$	0.039	0.102	0.000	1.431	54
17	Presence of internal star $_{t-4\ to\ t-1}$	0.542	0.499	0.000	1.000	32
18	Firm research quality $_{t-4\ to\ t-1}$	309.672	62.043	0.000	345.179	59
19	Firm research diversity $_{t-4\ to\ t-1}$	0.280	0.076	0.000	0.372	59

Table 2.1. continued.

ACADEMIC STAR CONTROLS						
20	Research quality of academic stars $_{t-4 \text{ to } t-1}$	94.582	65.065	0.000	276.200	55
21	Research diversity of academic stars $_{t-4 \text{ to } t-1}$	0.426	0.213	0.000	0.732	55
STAR-FIRM COLLABORATION CONTROLS						
22	quality of basic collaboration with academic stars $_{t-4 \text{ to } t-1}$	0.641	0.608	0.000	4.355	53
23	Collaborative experience firm and academic stars $_{t-4 \text{ to } t-1}$	0.600	1.807	0.000	17.308	46
24	Pseudo fixed effect	4.698	1.917	0.000	8.441	59

In order to identify the relationships under study limiting multi-collinearity concerns, while controlling for other key research characteristics of the firm, we employ a cascaded model setup. Firms' resources spent on R&D, which should have a direct relationship with innovation output, enter in absolute terms. The impact of in-house basic research is subsequently included as the ratio of the number of basic research publications in the previous four years over R&D expenditures, while the general effect of performing such basic research jointly with academia is measured by the ratio of collaborative to all basic research publications of the firm. The focal variables indicating academic star-firm collaboration under various contingencies are subsequently included as shares of basic research collaborations with academia (as illustrated by Figure 1). Hence, collaboration with academic non-star scientists is the reference group and the focal variables on collaboration types measure the 'innovation premium' of collaborating with stars under various contingencies rather than with other academic researchers.

Control Variables

The models include a range of control variables at the firm, academic star, and firm-star collaboration level to isolate the effects of star collaboration and its contingencies. Besides R&D, basic research, and collaboration with academia in basic research, we include a control for the technological diversity of a firm's technology portfolio, which in prior research has been shown to relate in a non-linear way to firms' innovation performance (e.g. Leten *et al.* 2007). Technological diversity is measured as the inverse Herfindahl index of the distribution of the four-year prior patent portfolio over 3-digit IPC patent classes. To allow for a non-linear relationship, this variable is included in both linear and quadratic form. The analysis also controls for the number of inter-firm research alliances during the past four years (taken from the SDC Platinum database), scaled by

R&D expenditures in the previous year (Owen-Smith and Walter W. Powell, 2004; Hess and Rothaermel, 2011; Belderbos, Gilsing and Suzuki, 2016). We include a dummy variable indicating if the firm employs an internal star scientist (Zucker *et al.* 2002; Hess and Rothaermel 2011).[19] We also include two additional academic research characteristics of the firm that may relate to innovation performance: the quality of the firm's research (measured by the average yearly number of citation-weighted publications in the previous four years), and the firm's research diversity (the

Blau index of the distribution of publications across the 44 scientific fields within Medicine and Life sciences during the previous four years). Two other control variables incorporates characteristics of the collaborating academic stars, to take into account possible residual heterogeneity among these extraordinary scientists. While our study focuses on partnerships with scientists in the upper tail of the quality and productivity distribution, there can still be star scientist heterogeneity influencing collaborative innovation performance outcomes. We include the research quality (the average citation-weighted publication count) and the research diversity of the collaborating stars' during the previous four years.

The analysis also includes two key characteristics at the star-firm collaboration level. Firms and stars may select the ex-ante most promising projects for collaboration and star scientists might be more willing to take on a specific collaboration configuration if the expected (scientific) impact of the project is promising. To control for this potential influence of heterogeneous project quality, the model includes the average forward citation rate of the collaborative publications. Second, new firm-star collaborations may be partially driven by prior collaboration experience, and such experience may have positive performance implications. Hence, the models include the average number of co-publications of the focal firm and the collaborating academic stars in the 4-year period prior to the period of the focal star-firm collaborations. Finally, in addition to the pseudo fixed firm effects, all models include year fixed effects to control for time specific shocks. Table 1 reports descriptive statistics. Almost all sample firms (55 out of 60) collaborated at least once with an academic star scientist during the sample period. Most of these firms (53) engaged in the most common type of academic star-firm collaboration: non-dedicated and non-translational collaboration with the academic star and not involving internal stars. Even the least frequent

[19] Including the natural logarithm of the number of internal star scientists did not deliver materially different results.

Table 2.2. Correlation matrix of all variables (406 obs.)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.
1.	1.00																						
2.	**0.70**	1.00																					
3.	-0.07	**-0.17**	1.00																				
4.	0.07	0.03	**0.13**	1.00																			
5.	**0.33**	**0.58**	0.08	**0.16**	1.00																		
6.	0.05	-0.10	**0.15**	**0.12**	**0.37**	1.00																	
7.	**0.28**	**0.38**	0.05	**0.13**	**0.52**	-0.03	1.00																
8.	**0.21**	**0.37**	-0.01	0.03	**0.44**	0.06	**0.11**	1.00															
9.	**0.13**	**0.44**	-0.04	**0.11**	**0.71**	-0.07	**0.21**	**0.14**	1.00														
10.	**0.27**	**0.31**	0.00	0.02	**0.29**	**0.12**	**0.10**	**0.19**	0.05	1.00													
11.	**0.34**	**0.37**	0.11	-0.07	**0.32**	0.00	**0.19**	0.06	0.06	**0.23**	1.00												
12.	0.06	**0.20**	**0.10**	-0.02	**0.35**	0.01	**0.15**	**0.29**	0.10	**0.20**	**0.23**	1.00											
13.	0.26	0.41	0.05	-0.00	**0.42**	0.00	**0.11**	**0.12**	**0.12**	**0.33**	**0.57**	**0.27**	1.00										
14.	**0.15**	**0.21**	0.03	-0.03	**0.13**	0.04	0.01	0.08	-0.04	**0.19**	**0.33**	**0.13**	**0.38**	1.00									
15.	**0.27**	**0.39**	**0.18**	**0.21**	**0.33**	0.02	**0.26**	**0.15**	**0.26**	-0.00	**0.22**	**0.10**	0.09	0.01	1.00								
16.	**-0.15**	**-0.32**	0.02	0.04	**-0.15**	**0.27**	**-0.16**	**-0.16**	**-0.22**	-0.06	-0.07	-0.07	-0.09	-0.07	**-0.10**	1.00							
17.	**0.45**	**0.68**	0.01	**-0.10**	**0.51**	0.01	**0.31**	**0.29**	**0.33**	**0.26**	**0.36**	**0.25**	**0.39**	**0.33**	**0.27**	**-0.23**	1.00						
18.	**0.14**	**0.33**	**0.15**	**0.52**	**0.30**	0.05	**0.18**	**0.15**	**0.27**	0.05	0.06	0.05	0.08	0.07	**0.31**	-0.08	**0.21**	1.00					
19.	**0.13**	**0.27**	**0.13**	**0.45**	**0.31**	0.06	0.06	**0.21**	**0.34**	0.02	-0.05	0.06	0.03	0.02	**0.22**	**-0.11**	**0.12**	**0.74**	1.00				
20.	**0.31**	**0.62**	0.08	0.07	**0.68**	-0.01	**0.34**	**0.25**	**0.62**	**0.25**	**0.27**	**0.30**	**0.43**	**0.12**	**0.36**	**-0.25**	**0.50**	**0.27**	**0.21**	1.00			
21.	**0.27**	**0.59**	0.09	**0.20**	**0.72**	0.07	**0.42**	**0.31**	**0.65**	**0.13**	**0.21**	**0.17**	**0.26**	**0.12**	**0.43**	**-0.27**	**0.49**	**0.39**	**0.33**	**0.76**	1.00		
22.	**0.36**	**0.53**	0.01	0.04	**0.50**	0.01	**0.26**	**0.17**	**0.43**	**0.23**	**0.25**	**0.17**	**0.32**	**0.11**	**0.29**	**-0.19**	**0.36**	**0.20**	**0.19**	**0.69**	**0.52**	1.00	
23.	**0.15**	**0.26**	0.06	0.03	**0.40**	0.06	0.09	**0.31**	**0.16**	**0.52**	**0.27**	**0.59**	**0.44**	**0.16**	0.08	-0.08	**0.24**	0.07	**0.15**	**0.37**	**0.19**	**0.26**	1.00
24.	**0.74**	**0.85**	0.06	**0.11**	**0.51**	-0.05	**0.38**	**0.32**	**0.32**	**0.29**	**0.39**	**0.21**	**0.39**	**0.22**	**0.49**	**-0.27**	**0.65**	**0.34**	**0.23**	**0.54**	**0.50**	**0.45**	**0.24**

Correlations in bold are significant at 0.05 level. Variables in grey colored cells are not jointly included in the empirical models.

collaboration type, dedicated and translational, is still practiced by 39 of the sample firms if this collaboration does not involve internal stars. In contrast, if internal stars are involved, only 13 firms have dedicated translational star collaborations. Of the 32 firms that employ internal star scientists, 29 involve them at least once in collaborations with academic stars.

Table 2 contains the correlation matrix. There is no evident multi-collinearity concern regarding the focal variables that are simultaneously included in the empirical models. A relatively high correlation is observed between R&D expenditures and the pseudo fixed effect, which is related to the relatively stable R&D budgets of firms over time.

2.4 EMPIRICAL RESULTS

Table 3 reports the regression results. Model 1 is the baseline model including only the control variables. In addition to the significant pseudo fixed effects ($p<0.0000$), firms that invest more in R&D ($p<0.000$), that collaborate more frequently with academia in basic research ($p=0.0010$. Firms' research quality and diversity show no significant association with innovation performance at conventional ($p<0.05$) significance levels. The academic star controls, research quality and research diversity have significantly negative and positive coefficients, respectively, but this significance disappears in the full specification of model 3. The quality of collaborative research with star scientists is positive and significantly associated with innovation performance ($p=0.0087$) in all models, while prior collaboration experience is also positive and gains significance in models 2 and 3.

Model 2 includes the share of co-publications in basic research with academic. The insignificant coefficient ($p=0.1405$) suggests that on average, star collaboration is not associated with an innovation performance premium over and above the positive influence of collaborating with academia. This overall effect warrants further investigation of the potential heterogeneities in the performance benefits of academic star scientist collaboration depending on collaboration characteristics, as hypothesized.

Model 3 includes the variables representing the shares of the eight specific collaboration types to examine the contingencies under which academic star scientist collaboration may be associated with a performance premium. There are marked differences between the coefficients of the shares

of different collaboration types. For basic research collaborations with academic stars but without internal stars (β_1), there is a positive coefficient (p=0.0016) for dedicated and translational collaboration (β_{1a}) and a negative coefficient (p=0.0026) for non-dedicated and non-translational collaboration (β_{1d}). Among the four shares of collaborations that involve both academic and internal star scientists (β_2), the only positive share (p=0.0280) is for dedicated and non-translational academic star collaboration (β_{2b}), while significantly negative coefficients (p<0.001) are observed for the two non-dedicated collaboration categories (β_{2c}, β_{2d}).

In order to test our hypotheses, we conduct chi-squared tests (reported in Table 4) that compare sets of coefficients for the collaboration shares. The first two hypotheses focus on the role of dedication and translation and are examined by comparing regression coefficients for the collaboration shares with and without dedication and translation, respectively. The sum of coefficients of the shares of dedicated academic star scientist collaboration is higher than the sum of coefficients of the shares of non-dedicated academic collaboration (chi2=12.14, p <0.001), indicating that there is a beneficial influence of dedication. This confirms hypothesis 1. A test comparing coefficients of collaboration shares with and without translational collaboration also indicates a difference (chi2=7.81, p=0.005), but translation has a weaker, rather than the hypothesized stronger relationship with innovation, rejecting hypothesis 2. Since we argue that the translational role of academic star scientists may be taken up by internal stars as well, we also examine if the role of translational activities by academic stars is positive for collaborations not involving internal stars. A test on the coefficients of collaboration shares with and without translation but not involving internal stars does show that translational collaboration has a stronger (chi2=5.82, p=0.016) positive association with innovation performance than non-translational collaboration. This result provides qualified support for hypothesis 2: there is a positive association between innovation and the degree to which academic star collaboration involves applied research, but only if these collaborations do not involve internal stars.

The final test in Table 4 relates to Hypothesis 3 and examines the consequences of involving internal star scientists in basic research collaborations with academic star scientists, depending on whether the academic star collaboration is translational or not. The test confirms that the association between innovation performance and joint internal and external star collaboration is

Table 2.3. Academic star collaboration in basic research and firms' innovation performance: Pseudo fixed effects Poisson panel regression results

		Model 1		Model 2		Model 3	
	β	Coef.	z-value (p-value)	Coef.	z-value (p-value)	Coef.	z-value (p-value)
KEY VARIABLES							
Log (R&D)		0.26	4.31 (0.0000)	0.27	4.54 (0.0000)	0.25	4.29 (0.0000)
Basic research publications / R&D		0.05	1.13 (0.2576)	0.06	1.25 (0.2129)	0.06	1.18 (0.2388)
Basic research co-publications with academia / Basic research publications		1.12	3.29 (0.0010)	1.12	3.30 (0.0010)	0.96	2.88 (0.0040)
Share of basic research co-publications with academia							
… without internal star and without academic stars		Reference group					
… with academic stars				-0.59	-1.47 (0.1405)		
… without internal stars and with academic stars							
… which are dedicated and translational	1a					1.96	3.15 (0.0016)
… which are dedicated and non-translational	1b					0.91	1.06 (0.2902)
… which are non-dedicated and translational	1c					0.32	0.30 (0.7675)
… which are non-dedicated and non-translational	1d					-2.03	-3.02 (0.0026)
… **with internal stars and with academic stars**							
… which are dedicated and translational	2a					-5.45	-1.50 (0.1345)
… which are dedicated and non-translational	2b					6.87	2.20 (0.0280)
… which are non-dedicated and translational	2c					-9.65	-4.03 (0.0001)
… which are non-dedicated and non-translational	2d					-3.84	-3.37 (0.0007)
… with internal stars and without academic stars						-0.36	-0.47 (0.6389)
FIRM CONTROLS							
Technological diversity		-0.03	-0.27 (0.7897)	-0.03	-0.30 (0.7651)	-0.02	-0.22 (0.8245)
Technological diversity ^{2}		-0.00	-0.36 (0.7166)	-0.00	-0.34 (0.7365)	-0.00	-0.45 (0.6540)
Research alliances / R&D		0.22	0.35 (0.7258)	0.22	0.34 (0.7371)	-0.08	-0.16 (0.8710)
Presence of internal star		-0.02	-0.16 (0.8746)	0.04	0.31 (0.7567)	0.06	0.50 (0.6161)
Firm research quality		-0.00	-1.92 (0.0555)	-0.00	-1.90 (0.0578)	-0.00	-1.91 (0.0566)
Firm research diversity		1.07	0.33 (0.7436)	1.08	0.33 (0.7411)	1.63	0.50 (0.6168)

Table 2.3 continued.

ACADEMIC STAR CONTROLS							
Research quality of academic stars		-0.00	-4.75 (0.0000)	-0.00	-4.19 (0.0000)	-0.00	-1.53 (0.1267)
Research diversity of academic stars		0.94	2.45 (0.0143)	1.05	2.78 (0.0055)	0.60	1.42 (0.1543)
STAR-FIRM COLLABORATION CONROLS							
Quality of basic collaborations with academic stars		0.18	2.62 (0.0087)	0.17	2.63 (0.0087)	0.18	2.88 (0.0039)
Collaborative experience of firm and academic stars		0.01	1.27 (0.2045)	0.02	1.98 (0.0477)	0.04	3.31 (0.0009)
Pseudo fixed effects		0.61	11.17 (0.0000)	0.61	11.16 (0.0000)	0.59	11.33 (0.0000)
Year dummies		Yes		Yes		Yes	
Constant		2.25	4.44 (0.0000)	2.24	4.44 (0.0000)	2.34	4.61 (0.0000)
Log likelihood		-31725.63		-31571.20		-27822.23	
LR test improved model fit (vs. Model 1)				308.86	(0.0000)	7806.80	(0.0000)

All models have 406 observations. P-value of robust standard errors in parentheses.

stronger (chi2=12.99, p=0.000) for non-translational than for translational collaboration, in support of the hypothesis.[20]

It is important to examine the economic significance of the estimation results. Exponentiated coefficients in the negative binomial model can be interpreted as pseudo-elasticities (incidence ratios). We infer that a standard deviation increase in the share of translational and dedicated star collaborations not involving internal stars is associated with a 13 percent increase in innovation performance. In contrast, increasing the share of non-translational and non-dedicated star collaborations by a standard deviation is associated with a 21 percent decrease in performance. For collaborations involving internal stars, the coefficients suggests a 7 percent increase (dedicated non-translational), and a decrease of 13 percent (non-dedicated translational and non-dedicated non-translational, respectively). However, these effects have to be assessed against the background of a simultaneous and sizeable positive effect of collaboration in basic research with academics, where a standard deviation increase is associated with a performance increase of 23 percent.

[20] We note that additional tests also showed that involving internal stars in translational academic star collaboration significantly decreases performance (chi2 = 12.81, p= 0.000). For collaborations with non-translational academic stars, involving internal stars only increases performance significantly when the academic stars are dedicated (chi2=3.04; p=0.081).

Is collaboration with academia in basic research by the sample firms overall positively associated with firm innovation performance? This question rises given the observed negative influences of specific types of academic star scientist collaboration, in conjunction with the positive influence of academic collaboration in basic research in general. Simulations at the firm level show that for about 8 percent of firm-year observations, the particular configuration of academic basic research collaboration is negatively associated with innovation performance, although this negative association is never statistically significant ($p < 0.01$). In contrast, for more than 73 percent of firm-year observations (297 out of 406) academic collaboration has a positive association with performance, while for 55 of these observations this association is statistically significant ($p < 0.01$). These findings suggest that while academic star scientist collaboration in basic research under unfavorable contingencies may lead to significantly negative performance premiums compared to non-star collaboration with academia, there is no evidence for an overall negative association between academic star scientist collaboration and firm innovation performance. At the same time, under favorable contingencies the results suggest clear positive effects of academic star scientist collaboration.

Table 2.4. Tests for the hypotheses on innovation performance benefits

	Chi2	Prob>chi2	Answer
Hypothesis 1: Are performance benefits greater in case of dedicated collaboration?			
$\beta1a + \beta1b + \beta2a + \beta2b - \beta1c - \beta1d - \beta2c - \beta2d = 0$ (1.96) + (0.91) + (-5.45) + (6.87) – (0.32) – (-2.03) – (-9.65) – (-3.84) = 19.49	12.14	0.001	Yes
Hypothesis 2: Are performance benefits greater in the case of translational collaboration?			
Across all star collaboration modes: $\beta1a + \beta1c + \beta2a + \beta2c - \beta1b - \beta1d - \beta2b - \beta2d = 0$ (1.96) + (0.32) + (-5.45) + (-9.65) – (0.91) – (-2.03) – (6.87) – (-3.84) = -14.73	7.81	0.005	No (reduced benefit)
Star collaboration modes not involving internal stars: $\beta1a + \beta1c - \beta1b - \beta1d = 0$ (1.96) + (0.32) – (0.91) – (-2.03) = 3.40	5.82	0.016	Yes
Hypothesis 3: Are performance benefits of involving internal stars greater when collaboration is non-translational?			
$\beta2b + \beta2d - \beta1b - \beta1d - (\beta2a + \beta2c - \beta1a - \beta1c) = 0$ (6.87) + (-3.84) – (0.91) – (-2.03) – (-5.45) – (-9.65) + (1.96) + (0.32) = 21.53	12.99	0.000	Yes

2.5 SUPPLEMENTARY ANALYSIS

In the following, the supplementary analysis performed to test the robustness of our results are discussed. The tables that present the results and descriptive statistics that will be discussed are included in appendix.

2.5.1 Partner matching

The emerging literature on matching in research collaboration has suggested that particular matches between the capabilities of a firm and its academic partner may be more productive (Mindruta 2013; Banal-Estanol 2017). These studies focused on the entire spectrum of firms and academics rather than on the very top of the distribution (star scientists). Nevertheless, it is worthwhile investigating whether the suggested performance-enhancing influences are observed in the context of our study. Based on these prior studies, we augment the model with a measure of positive assortative matching on research quality and basic research orientation (Banal-Estanol *et al.* 2017) and negative assortative matching on research diversity (Mindruta 2013). The matching variables are constructed by assessing, for each star-firm collaboration, how the star (and firm) scores relative to the mean for all stars (firms) on these dimensions. If a star and firm score similar (i.e. both are above or below the mean) on a dimension, the star-firm collaboration is classified as positive assortative matching for this dimension, and vice versa for negative assortative matching. The matching variables measure the share of star collaborations for which there is positive assortative matching (research quality) and negative assortative matching (research diversity) between the firm and the star.

The results are shown in Table A1. The empirical results do not suggest a further influence of matching, while the focal results and hypothesis tests remain unchanged. This is likely to be related to our focused and selective context of star scientist collaboration in the pharmaceutical industry and the fact that the analysis already controls for collaboration quality and experience.

2.5.2 Heterogeneity in the scientific quality of types of collaboration projects

We examine more in detail whether the choice for certain types of academic star scientist collaborations may be associated with the underlying quality of the proposed research collaborations. While our main models control for the average quality of collaborative basic research with academic star scientists, we also explore an expansion of the model in which we include eight variables that capture the quality of the projects for each type of star scientist

collaboration, measured by the forward citations to the associated co-authored papers. Table A2 shows that variation in collaboration type-specific research quality explains some residual variation in innovation benefits for non-dedicated collaborations involving internal stars, while there appears to be a negative association between research quality and innovation for non-dedicated non-translation collaboration without internal star involvement. The focal results of our analysis do not change, with Hypotheses 1 and 3 supported, while support for Hypothesis 2, as before, is confined to collaborations without internal star involvement.

2.5.3 Heterogeneity experience across types of collaboration projects

We examined whether there is heterogeneity in the influence of past collaboration experience depending on the type of collaboration. Table A3 reports the empirical results if eight experience variables are included, one for each academic star collaboration type. The results suggest that experience has a significant association with innovation for translational collaborations without internal stars, indicating that experience can enhance the performance benefits of star collaborations that involve both basic and applied research. The focal results and tests for our hypotheses are not affected, but we do note that the test statistic for a performance premium of translational star without internal star involvement becomes insignificant at conventional levels, which likely reflects a lesser ability to estimate all coefficients with precision. Note that the result for dedication remains intact after introducing the detailed controls for experience, suggesting that - once we account for an established trust relationship between the firm and the star - the dedicated nature of the collaboration primarily serves to avoid unwanted knowledge spillovers.

2.5.4 Extension of the patent citation window

We examined whether the results are sensitive to the choice of the citation window for the innovation performance measure. Employing an extended eight-year citation window, we observed no substantive differences in the empirical results, as shown in Table A4.

2.5.5 Firm heterogeneity and reverse causality

It is conceivable that a positive correlation between the focal collaboration variables and firms' innovation performance may also be related to the ability of top-performing firms to attract and persuade academic star scientists to engage in (specific forms of) collaboration. We examine whether there is evidence of such reverse causality by examining the relationship between firms' engagement in the specific star collaboration types and their prior innovation performance. The

dependent variables are the core independent variables in the focal model in the paper: the shares of the specific types of academic basic research collaboration of a firm in a 4-year window. Since we have R&D data from 1994 and collaboration data up to 2001, this leaves a maximum of four observations on the dependent variable per firm (1995-1998, 1996-1999, 1997-2000, 1998-2001) for a total of 179 observations.

We use a fractional response model that allows for simultaneous estimation of star-firm collaboration shares and the inclusion of pseudo fixed effects (pre-sample observations on collaboration shares). Table A5 shows the results of regressing the shares of specific types of basic research collaboration that firms engage in on firms' past innovation performance and the set of control variables that is also included in the main analysis in the paper. We observe no significantly positive relationships between past patent performance and engagement in those collaboration types that the main analysis found to be associated with positive performance effects. Past innovation appears to be negatively associated with dedicated non-translational collaboration without internal stars, and positively associated with non-dedicated translational collaboration with internal stars. The findings do not suggest that our core results on the importance of dedication and translation are biased due to reverse causality. In this regard, the findings are in line with the matching analysis of Mindruta (2013) suggesting that high patent capabilities of firms are not a precursor of science collaboration.

2.5.6 Star heterogeneity and translation, dedication

We additionally performed a supplementary analysis to examine whether there is heterogeneity among the stars related to their mode of collaboration. For this robustness check, we focused on academic star scientists that have at least one basic research collaboration with our sample firms (6 554 scientists). The probability that these stars will behave dedicated or/and translational in at least one of these collaborations is investigated by Probit regressions corrected for heteroscedasticity. While Table A6 presents the variables with their description and descriptive statistics, Table A7 shows the regression results for the probability to collaborate in a dedicated manner, a translational manner and both manners within the same collaboration. It is noticeable that the profile of a dedicated star differs considerably from that of a translational academic star. Although both dedicated and translational behavior is more likely for young academic stars, dedication is mainly stimulated by more last author positions and a more specialized research

profile while translation is mainly stimulated by more first author positions and a more applied research focus. Finally, the combination of dedicated and translational behavior in a collaboration seems to be stimulated by less first-author positions and a more applied research profile. Although, multi-disciplinarily and higher research output both in publication count and citation rate also have a positive and significant influence on this behavior, their average marginal effects are negligible. Overall, we must acknowledge that all the average marginal effects are limited considering the standard deviation of the dependent variables. Hence, the investigated star characteristics will only have a limited influence on the results of our study. Especially, heterogeneity based on research output appears not to have driven our results as the related variables have a negative impact on dedicated behavior and only a very limited positive influence on translation.

2.6 CONCLUSION AND DISCUSSION

Our analysis of the relationship between pharmaceutical firms' innovation performance and their basic research collaborations with academic star scientists revealed that such collaborations on average are not associated with a performance premium compared to academic collaborations not involving stars. Working with the best and brightest academic scientists does not only provide advantages to the firm, but may face important obstacles related to the full agenda and stronger independence of the star - which may lead to reduced commitment and interaction -, a lesser emphasis on translation, and undesired knowledge spillovers to other firms.

When firms hope to increase their innovation performance by collaboration with academia, this study indicates that collaborations with academic star scientists may only represent a beneficial share within this portfolio of collaboration when managed under the right conditions. Although firms may not always be in a position to demand such conditions, we find that dedication and translation jointly are key conditions to observe an innovation performance premium of academic star collaboration in comparison with non-star collaboration. Dedication (the academic star abstains from collaboration with other firms) promotes commitment and trust and reduces the threat of knowledge dissipation to other firms. Translation (the collaboration with the academic star extends to applied research) is important to fruitfully apply insights from basic research to applied research trajectories.

If the basic research collaboration with an academic star also involves an internal star scientist, applied research involvement of the academic star is not required, and can even turn out to be detrimental for a performance premium, presumably because the internal star can fulfill this translational role. Only one configuration of basic research collaboration involving both internal and academic star scientists is associated with a performance premium, namely a collaboration in which academic stars are dedicated but not involved in applied collaborative research. The lack of more general innovation performance benefits in case of joint involvement of internal and academic stars may indicate problems related to knowledge redundancy, hierarchy and conflicts, with non-dedicated collaboration furthermore increasing the risk of knowledge dissipation. Overall, the observed patterns indicate that dedication is the most crucial condition to safeguard the innovative return of collaborations in basic research with academic star scientists, with positive influences regardless of the involvement of internal stars.

Our study emphasizes the difficulty of transferring and safeguarding knowledge in the context of university-industry collaboration, the connected roles of basic and applied research, and the specific role of internal star scientists for innovation performance. In doing so, our research contributes to both the broad stream of research on collaboration between firms and academia and the literature on the role of in-house star scientists. We contribute to the literature on industry-science collaboration by suggesting the importance of dedication and translation as crucial contingencies for firms to achieve an innovation premium when collaborating with academic star scientists in basic research. The importance of *dedication* has implications for our understanding of the organization of knowledge networks that span the institutional boundary between academia and industry. More specifically, linkages among organizations have been conceptualized on a continuum between 'closed conduits' where only the linked entities benefit from the exchanged knowledge, and 'porous channels' that allow for knowledge spillovers to external entities (Owen-Smith & Powell, 2004). We find that dedicated collaborations with academic star scientists – reflecting the aforementioned 'closed conduits' rather than 'porous channels' – result in an innovation advantage for firms. The importance of *translation* is consistent with the notion in the literature of the role of 'Pasteur' or 'bridging' scientists (Gittelman and Kogut 2003; Subramanian, Lim, and Soh 2013; Baba, Shichijo, and Sedita 2009) as important collaboration partners of firms. However, our study suggests an importance nuance, as the role of translational academic star

scientists depends on whether such translational capabilities are available among collaborating in-house star scientists.

We contribute to the debate on the role of internal star scientists by highlighting that when they are involved in collaborations with academic stars, a specific configuration of the collaboration is required to bring out benefits. Our study provides a first analysis of the interactive roles of internal and academic star scientists, which have been examined only separately in prior work (e.g. Lynne G. Zucker, Darby and Armstrong, 2002b; Rothaermel and Hess, 2007; Hess and Rothaermel, 2011, 2012; P. Almeida, Hohberger and Parada, 2011; Grigoriou and Rothaermel, 2013; Tzabbar and Kehoe, 2013; Kehoe and Tzabbar, 2015). Our findings show that this interplay is of importance, as the internal star scientist seems able, and is possibly even more capable, to take up the translational role within an academic star-firm collaboration in basic research. Our findings on the one hand confirm the notion that internal and external star scientists can be considered as substitutive sources of knowledge (Hess and Rothaermel 2011; Subramanian *et al.* 2013), leading to underperformance of collaboration. On the other hand, we highlight that this substitution effect can be mitigated and positive collaboration premiums can be observed if there is a proper task division in collaboration with the internal star focusing on translation.

In closing, we note the most salient limitations of our study. First, our study focuses on large, vertically integrated pharmaceutical firms, and is not representative for smaller, specialized (biotechnology) firms for which ties to academic star scientists may be of a different nature due to the closeness of the biotech and university communities (Powell, Koput, and Smith-Doerr 1996). Our findings for the pharmaceutical industry may neither be representative of basic research collaborations with star scientists in other science-based industries such as the electronics industry where new and existing technologies complement each other in complex interrelations. Second, although we used an elaborate specification to ensure that our findings are not confounded by selection bias, we cannot fully exclude the possibility of additional heterogeneity in projects and the related 'matching' between firms, star scientists and collaborative contingencies. Third, our analysis focused on the individual and did not explicitly examine team effects: star scientists often manage laboratories and employ other scientists who may take up specific collaboration tasks. Given the difficulties in characterizing a very large group of individual non-star co-authors, our analysis did not examine potential comparable influences of collaboration types for basic research

not involving star scientists. Finally, our analysis only focused on the association between star-firm collaboration and the firm's innovation performance. Collaborations may however also have other objectives such as signaling research competence to potential employees and general reputation building. Hence, even if specific types of star collaboration are negatively associated with direct innovation success, longer term effects may compensate, and the motivation for star collaborations may lie elsewhere. Motivations for, and implications of, basic research collaboration and in particular dedication and translation will also be different if examined from the perspective of the academic star scientist. We hope that future research can provide more systematic insights into the variety of drivers of basic research collaborations, in particular from the perspective of the (star) scientists. The limitations of our study suggest a rich agenda for future research.

2.7 APPENDIX

Table 2.A1. Academic star collaboration in basic research and firms' innovation performance: firm-star matching

	β	Model 1		Model 2		Model 3	
		Coef.	z-value (p-value)	Coef.	z-value (p-value)	Coef.	z-value (p-value)
KEY VARIABLES							
Log (R&D)		0.26	4.25 (0.0000)	0.26	4.44 (0.0000)	0.24	4.02 (0.0001)
Basic research publications / R&D		0.04	0.87 (0.3849)	0.05	0.96 (0.3357)	0.04	0.87 (0.3833)
Basic research co-publications with academia / Basic research publications		1.10	3.39 (0.0007)	1.10	3.41 (0.0006)	0.94	2.88 (0.0039)
Share of basic research co-publications with academia							
… without internal star and without academic stars		Reference group					
… with academic stars				-0.48	-1.16 (0.2468)		
… without internal stars and with academic stars							
… which are dedicated and translational	1a					2.09	3.30 (0.0010)
… which are dedicated and non-translational	1b					1.53	1.67 (0.0947)
… which are non-dedicated and translational	1c					0.35	0.33 (0.7411)
… which are non-dedicated and non-translational	1d					-1.99	-2.94 (0.0033)
… **with internal stars and with academic stars**							
… which are dedicated and translational	2a					-5.34	-1.60 (0.1088)
… which are dedicated and non-translational	2b					6.06	1.97 (0.0487)
… which are non-dedicated and translational	2c					-9.73	-4.25 (0.0000)
… which are non-dedicated and non-translational	2d					-3.81	-3.46 (0.0005)
… with internal stars and without academic stars						-0.10	-0.14 (0.8884)
FIRM CONTROLS							
Technological diversity		-0.03	-0.30 (0.7629)	-0.03	-0.32 (0.7454)	-0.02	-0.23 (0.8177)
Technological diversity ^{2}		-0.00	-0.33 (0.7405)	-0.00	-0.31 (0.7559)	-0.00	-0.47 (0.6404)
Research alliances / R&D		0.11	0.16 (0.8727)	0.12	0.17 (0.8631)	-0.26	-0.46 (0.6473)
Presence of internal star		0.00	0.01 (0.9924)	0.04	0.37 (0.7082)	0.06	0.48 (0.6287)
Firm research quality		-0.01	-1.97 (0.0494)	-0.01	-1.99 (0.0471)	-0.00	-1.43 (0.1528)
Firm research diversity		1.95	0.59 (0.5574)	2.01	0.60 (0.5467)	0.80	0.24 (0.8082)

Table 2.A1 continued.

STAR CONTROLS							
research quality of academic stars		-0.01	-4.72 (0.0000)	-0.01	-4.38 (0.0000)	-0.00	-2.26 (0.0237)
research diversity of academic stars		1.92	2.34 (0.0193)	1.88	2.31 (0.0211)	1.75	2.25 (0.0245)
STAR-FIRM COLLABORATION CONTROLS							
quality of basic collaborations with academic stars		0.21	3.05 (0.0023)	0.21	3.02 (0.0026)	0.22	3.03 (0.0025)
experience between firm and academic stars		0.01	1.58 (0.1139)	0.02	2.08 (0.0373)	0.04	3.45 (0.0006)
STAR-FIRM MATCHING CONTROLS							
Share of basic coll. with academic stars							
… assortative matching on research quality		-0.84	-2.01 (0.0449)	-0.84	-2.05 (0.0407)	-0.46	-1.17 (0.2433)
… assortative matching on basic research focus		-0.95	-0.94 (0.3481)	-0.88	-0.87 (0.3831)	0.38	0.38 (0.7037)
… anti-assortative matching on research diversity		0.93	0.77 (0.4419)	0.99	0.82 (0.4127)	-1.14	-0.95 (0.3405)
Pseudo fixed effects		0.62	11.31 (0.0000)	0.61	11.30 (0.0000)	0.60	11.24 (0.0000)
Year dummies		Yes		Yes		Yes	
Constant		2.24	4.63 (0.0000)	2.24	4.62 (0.0000)	2.36	4.78 (0.0000)
Log likelihood		-31725.63		-31571.20		-27822.23	
LR test improved model fit (vs. Model 1)				188.11	(0.0000)	7294.51	(0.0000)
HYPOTHESES TESTS							
		CHI^2		($Prob > CHI^2$)		Answer	
H1		13.61		(0.000)		Yes	
H2 all collaborations		8.23		(0.004)		No (reduced benefit)	
collaborations without internal stars		4.22		(0.040)		Yes	
H3		12.33		(0.000)		Yes	

Notes: All models have 406 observations. P-value of robust standard errors in parentheses.

Table 2.A2. Academic star collaboration in basic research and firms' innovation performance: collaborative research quality heterogeneity

	β	Coef.	z-value (p-value)
KEY VARIABLES			
Log (R&D)		0.26	4.54 (0.0000)
Basic research publications / R&D		0.05	0.94 (0.3476)
Basic research co-publications with academia / Basic research publications		0.91	2.74 (0.0062)
Share of basic research co-publications with academia			
… without internal star and without academic stars		Reference group	
… without internal stars and with academic stars			
… which are dedicated and translational	1a	2.09	3.51 (0.0004)
… which are dedicated and non-translational	1b	0.88	1.03 (0.3024)
… which are non-dedicated and translational	1c	0.34	0.33 (0.7387)
… which are non-dedicated and non-translational	1d	-1.99	-2.82 (0.0048)
…**with internal stars and with academic stars**			
… which are dedicated and translational	2a	-4.21	-0.96 (0.3374)
… which are dedicated and non-translational	2b	7.35	2.35 (0.0189)
… which are non-dedicated and translational	2c	-9.09	-4.24 (0.0000)
… which are non-dedicated and non-translational	2d	-4.02	-3.46 (0.0005)
…with internal stars and without academic stars		-0.19	-0.27 (0.7882)
FIRM CONTROLS			
Technological diversity		-0.01	-0.13 (0.8946)
Technological diversity ^{2}		-0.00	-0.51 (0.6090)
Research alliances / R&D		0.00	0.00 (0.9996)
Presence of internal star)		0.02	0.16 (0.8734)
Firm research quality		-0.00	-1.87 (0.0619)
Firm research diversity		1.60	0.49 (0.6267)

Table 2.A2 continued.

STAR CONTROLS				
research quality of academic stars		-0.00		-0.86 (0.3924)
research diversity of academic stars		0.55		1.31 (0.1912)
STAR-FIRM COLLABORATION CONTROLS				
experience between firm and academic stars		0.02		0.38 (0.7008)
quality of				
… collaborations without internal stars, with dedicated and translational academic stars		-0.00		-0.14 (0.8874)
… collaborations without internal stars, with dedicated and non-translational academic stars		0.01		0.71 (0.4783)
… collaborations without internal stars, with non-dedicated and translational academic stars		0.00		0.03 (0.9794)
… collaborations without internal stars, with non-dedicated and non-translational academic stars		-0.19		-2.15 (0.0314)
… collaborations with internal stars, with dedicated and translational academic stars		0.04		0.91 (0.3603)
… collaborations with internal stars, with dedicated and non-translational academic stars		0.04		0.95 (0.3445)
… collaborations with internal stars, with non-dedicated and translational academic stars		0.09		1.66 (0.0959)
… collaborations with internal stars, with non-dedicated and non-translational academic stars		0.04		2.98 (0.0029)
Pseudo fixed effects		0.60		11.05 (0.0000)
Year dummies		Yes		
Constant		2.25		4.49 (0.0000)
Log likelihood		-27504.08		
HYPOTHESES TESTS				
		CHI^2	(Prob > CHI^2)	Answer
H1		11.90	(0.001)	Yes
H2 all collaborations		5.67	(0.017)	No (reduced benefit)
collaborations without internal stars		7.28	(0.007)	Yes
H3		10.79	(0.001)	Yes

Notes: All models have 406 observations. P-value of robust standard errors in parentheses.

Table 2.A3. Academic star collaboration in basic research and firms' innovation performance: collaborative research experience heterogeneity

	β	Coef.	z-value (p-value)
KEY VARIABLES			
Log (R&D)		0.23	3.55 (0.0004)
Basic research publications / R&D		0.04	0.86 (0.3883)
Basic research co-publications with academia / Basic research publications		0.86	2.60 (0.0094)
Share of basic research co-publications with academia			
… without internal star and without academic stars		Reference group	
… without internal stars and with academic stars			
… which are dedicated and translational	1a	1.79	2.74 (0.0061)
… which are dedicated and non-translational	1b	1.04	1.17 (0.2404)
… which are non-dedicated and translational	1c	-0.19	-0.21 (0.8359)
… which are non-dedicated and non-translational	1d	-1.69	-2.49 (0.0128)
…**with internal stars and with academic stars**			
… which are dedicated and translational	2a	-5.35	-1.35 (0.1772)
… which are dedicated and non-translational	2b	8.57	2.85 (0.0044)
… which are non-dedicated and translational	2c	-8.85	-3.09 (0.0020)
… which are non-dedicated and non-translational	2d	-3.24	-2.48 (0.0132)
…with internal stars and without academic stars		-0.47	-0.63 (0.5301)
FIRM CONTROLS			
Technological diversity		-0.03	-0.35 (0.7294)
Technological diversity ^{2}		-0.00	-0.32 (0.7453)
Research alliances / R&D		-0.09	-0.17 (0.8619)
Presence of internal star)		0.04	0.33 (0.7427)
Firm research quality		-0.00	-1.82 (0.0694)
Firm research diversity		1.55	0.49 (0.6253)

Table 2.A3 continued.

STAR CONTROLS				
research quality of academic stars		-0.00		-1.96 (0.0497)
research diversity of academic stars		0.61		1.43 (0.1531)
STAR-FIRM COLLABORATION CONTROLS				
quality of basic collaborations with academic stars		0.18		2.91 (0.0037)
experience co-publications between firm and academic stars of …				
… collaborations without internal stars, with dedicated and translational academic stars		0.02		1.85 (0.0641)
… collaborations without internal stars, with dedicated and non-translational academic stars		0.06		0.80 (0.4209)
… collaborations without internal stars, with non-dedicated and translational academic stars		0.04		1.67 (0.0957)
… collaborations without internal stars, with non-dedicated and non-translational academic stars		-0.04		-0.93 (0.3546)
… collaborations with internal stars, with dedicated and translational academic stars		-0.01		-0.96 (0.3352)
… collaborations with internal stars, with dedicated and non-translational academic stars		-0.03		-1.01 (0.3121)
… collaborations with internal stars, with non-dedicated and translational academic stars		-0.02		-1.30 (0.1922)
… collaborations with internal stars, with non-dedicated and non-translational academic stars		0.01		0.33 (0.7440)
Pseudo fixed effects		0.62		11.91 (0.0000)
Year dummies		Yes		
Constant		2.37		4.73 (0.0000)
Log likelihood		-26811.10		
HYPOTHESES TESTS				
		CHI^2	(Prob > CHI^2)	Answer
H1		11.22	(0.001)	Yes
H2 all collaborations		8.18	(0.004)	No (reduced benefit)
collaborations without internal stars		2.64	(0.1041)	No
H3		12.04	(0.001)	Yes

Notes: All models have 406 observations. P-value of robust standard errors in parentheses.

Table 2.A4. Academic star collaboration in basic research and firms' innovation performance: 8-year citation window for the dependent variable

	β	Coef.	z-value (p-value)
KEY VARIABLES			
Log (R&D)		0.20	3.44 (0.0006)
Basic research publications / R&D		0.06	1.19 (0.2357)
Basic research co-publications with academia / Basic research publications		0.72	2.16 (0.0311)
Share of basic research co-publications with academia			
… without internal star and without academic stars		Reference group	
… without internal stars and with academic stars			
… which are dedicated and translational	1a	2.16	2.92 (0.0035)
… which are dedicated and non-translational	1b	0.90	0.96 (0.3348)
… which are non-dedicated and translational	1c	0.23	0.23 (0.8196)
… which are non-dedicated and non-translational	1d	-2.14	-2.98 (0.0029)
…**with internal stars and with academic stars**			
… which are dedicated and translational	2a	-3.70	-1.08 (0.2785)
… which are dedicated and non-translational	2b	6.51	2.06 (0.0395)
… which are non-dedicated and translational	2c	-8.97	-3.73 (0.0002)
… which are non-dedicated and non-translational	2d	-3.30	-2.91 (0.0037)
…with internal stars and without academic stars		-1.00	-1.25 (0.2102)
FIRM CONTROLS			
Technological diversity		-0.05	-0.48 (0.6326)
Technological diversity ^{2}		-0.00	-0.18 (0.8539)
Research alliances / R&D		-0.48	-0.79 (0.4292)
Dummy (presence of internal star)		0.06	0.52 (0.6026)
Firm research quality		-0.00	-1.67 (0.0953)
Firm research diversity		1.09	0.37 (0.7089)

Table 2.A4 continued.

STARS CONTROLS			
research quality of academic stars		-0.00	-1.49 (0.1350)
research diversity of academic stars		0.59	1.45 (0.1465)
STAR-FIRM COLLABORATION CONTROLS			
experience between firm and academic stars		0.17	2.66 (0.0079)
quality of basic collaborations with academic stars		0.04	3.30 (0.0010)
Pseudo fixed effects		0.64	12.63 (0.0000)
Year dummies		Yes	
Constant		2.63	5.69 (0.0000)
Log likelihood		-27504.08	
HYPOTHESES TESTS			
		CHI^2 (Prob > CHI^2)	Answer
H1		13.20 (0.000)	Yes
H2 all collaborations		5.75 (0.017)	No (reduced benefit)
collaborations without internal stars		6.92 (0.009)	Yes
H3		11.04 (0.001)	Yes

Notes: All models have 406 observations. P-value of robust standard errors in parentheses.

Table 2.A5. Reverse causality? Firms' prior patent performance and engagement in different collaboration types

	Without internal stars				With internal stars			
	With academic stars which are				With academic stars which are			
	dedic. and transl.	*dedic.and non-transl.*	*non-dedic. and transl.*	*non-dedic., non-transl.*	*dedic. and transl.*	*dedic. and non-transl.*	*non-dedic. and transl.*	*non-dedic., non-transl.*
Mean value	0.0109	0.0498	0.0331	0.1581	0.0019	0.0027	0.0051	0.0135
Explanatory variables:	Marginal Effects							
Citation weighted patent stock $_{\text{from t-1 to t-4}}$ / 1000	0.00 (0.3462)	-0.01 (0.0066)	-0.00 (0.8209)	-0.00 (0.2349)	-0.00 (0.9667)	0.00 (0.5691)	0.00 (0.0001)	-0.00 (0.2897)
Log(R&D $_{\text{t-1, in million USD}}$)	0.00 (0.6049)	0.00 (0.6612)	0.01 (0.1653)	0.01 (0.0012)	0.00 (0.1430)	0.00 (0.1209)	0.00 (0.6032)	0.00 (0.0452)
Technological diversity $_{\text{t-4 to t-1, IPC3}}$	0.00 (0.1837)	0.02 (0.0950)	0.02 (0.0019)	0.02 (0.0006)	-0.00 (0.0735)	-0.00 (0.5188)	-0.00 (0.7083)	0.00 (0.1321)
Technological diversity^2 $_{\text{t-4 to t-1, IPC3}}$	-0.00 (0.1909)	-0.00 (0.1779)	-0.00 (0.0116)	-0.00 (0.0001)	0.00 (0.1365)	-0.00 (0.8496)	-0.00 (0.8718)	-0.00 (0.1296)
Research alliances $_{\text{t-4 to t-1, in million USD}}$ / R&D $_{\text{t-1, in million USD}}$	0.01 (0.5843)	0.26 (0.0001)	-0.02 (0.6001)	-0.02 (0.6267)	0.00 (0.6142)	0.03 (0.1334)	0.01 (0.2027)	0.01 (0.6007)
Dummy(presence of internal star$_{\text{t-4 to t-1}}$)	0.00 (0.5495)	0.00 (0.8267)	-0.00 (0.9373)	-0.02 (0.0013)	0.00 (0.1926)	0.00 (0.4328)	0.00 (0.3557)	0.01 (0.1678)
Basic research publications $_{\text{t-4 to t-1}}$ / R&D $_{\text{t-1, in million USD}}$	-0.00 (0.1338)	-0.01 (0.2046)	-0.00 (0.6032)	0.00 (0.9700)	0.00 (0.7270)	0.01 (0.0082)	0.00 (0.0003)	0.01 (0.0003)
Basic co-publications with academia $_{\text{t-4 to t-1}}$ / Basic research publications $_{\text{t-4 to t-1}}$	-0.01 (0.1866)	-0.14 (0.0000)	-0.03 (0.0621)	-0.02 (0.2001)	0.00 (0.6078)	0.01 (0.4290)	0.00 (0.7353)	-0.01 (0.2445)
Average scientific quality of basic coll. with academic stars $_{\text{t-4 to t-1}}$	-0.00 (0.1034)	0.04 (0.0001)	0.00 (0.4341)	-0.02 (0.0001)	0.00 (0.6434)	0.01 (0.0458)	0.00 (0.4571)	-0.00 (0.4970)
Pre-sample Value (pseudo fixed effect)	0.09 (0.1574)	0.14 (0.3204)	-0.10 (0.2876)	-0.01 (0.7975)	0.38 (0.0004)	-0.15 (0.3495)	0.14 (0.0195)	0.00 (0.9337)
Pre-sample Dummy (1 if DV = 0 in pre-sample period)	0.00 (0.5404)	-0.06 (0.0049)	0.00 (0.8661)	-0.03 (0.0149)	0.00 (0.3334)	-0.03 (0.0001)	-0.00 (0.3837)	0.01 (0.0353)

Results of simultaneous fractional response analysis with pseudo fixed effects. p-value in parentheses. Year dummies included. 179 firm-year observations.

Table 2.A6. Investigating star heterogeneity: description and descriptive statistics of the variables

	Description	**Mean**	**Std.**
Dependent Variables			
Dummy dedication	1 if acad. star behaves dedicated in at least one basic-research collaboration with a sample firm from t-1 to t-4	0.507	0.500
Dummy translation	1 if acad. star behaves translational in at least one basic-research collaboration with a sample firm from t-1 to t-4	0.229	0.420
Dummy dedication and translation	1 if acad. star behaves both dedicated and translational in at least one basic-research collaboration with a sample firm from t-1 to t-4	0.057	0.233
Independent Variables			
Career age	Number of years between star's first publication in Pubmed and year t	23.063	9.615
Share of publications authored in first position	Number of publications from t-1 to t-4 with the acad. star as first author/ total number of star's publications from t-1 to t-4	0.433	0.234
Share of publications authored in last position	Number of publications from t-1 to t-4 with the acad. star as last author/ total number of star's publications from t-1 to t-4	0.131	0.124
Applied-research focus	Number of star's publications from t-1 to t-4 in applied research / total number of star's publications from t-1 to t-4	0.300	0.274
Normalized publication count	Normalized number of star's publications from t-1 to t-4 averaged across fields in which the acad. star is identified as a star in year t	0.726	0.778
Normalized citation rate	Normalized number of star's citations per publication from t-1 to t-4 averaged across fields in which the acad. star is identified as a star in year t	7.142	4.118
Multidisciplinary profile	Number of fields in which the acad. star is identified as a star in year t	1.462	0.782

*16,447 observations considering for every academic star each year t for which she has at least one basic-research collaboration with a sample form from t-1 to t-4

Table 2.A7. Investigating star heterogeneity: Probit regression results robust for heteroscedasticity

Dependent Variable	**Dummy dedication**		**Dummy translation**		**Dummy dedication and translation**	
	AME	(p-value)	AME	(p-value)	AME	(p-value)
Career age	-0.002	(0.005)	-0.002	(0.007)	-0.000	(0.171)
Share of publications authored in first position	0.074	(0.000)	0.133	(0.003)	-0.011	(0.001)
Share of publications authored in last position	0.295	(0.000)	0.063	(0.104)	-0.002	(0.267)
Applied-research focus	-0.027	(0.034)	0.598	(0.000)	0.005	(0.000)
Normalized publication count	-0.079	(0.000)	0.031	(0.868)	0.000	(0.000)
Normalized citation rate	-0.024	(0.000)	0.010	(0.000)	-0.001	(0.000)
Multidisciplinary profile	-0.145	(0.000)	0.033	(0.000)	0.000	(0.000)
Year dummies	Included		Included		Included	
Constant		(0.000)		(0.000)		(0.000)
Number of observations		16,447		16,447		16,447

Table 2.A8. Firms included in the sample

Company Name	R&D expenditures in 2002 in million US dollars	Head Quarter Location	Company Name	R&D expenditures in 2002 in million US dollars	Head Quarter Location
ABBOTT LABORATORIES	1,561	US	MOCHIDA PHARMACEUTICAL	69	JP
AJINOMOTO	215	JP	MYLAN LABORATORIES	58	US
ALLERGAN	233	US	NOVARTIS	278	EU
ALTANA	347	EU	NOVO NORDISK AS	499	EU
ASTRAZENECA	3,047	EU	ONO PHARMACEUTICAL	242	JP
AVENTIS	3,218	EU	PFIZER	5,175	US
BARR LABORATORIES	75	US	PLIVA	64	EU
BECTON DICKINSON	220	US	RECORDATI	33	EU
BOEHRINGER INGELHEIM	1,227	EU	ROCHE	3,022	EU
BRISTOL MYERS SQUIBB	2,217	US	SANKYO	691	JP
DADE BEHRING	28	US	SANOFI SYNTHELABO	1,146	EU
DAIICHI PHARMACEUTICAL	440	JP	SANTEN PHARMACEUTICAL	101	JP
DAINIPPON PHARMACEUTICAL	121	JP	SCHERING	891	EU
EGIS PHARMACEUTICALS	194	EU	SCHERING PLOUGH	1,424	US
EISAI	476	JP	SCHWARZ PHARMA	116	EU
ELI LILLY	2,149	US	SEIKAGAKU	29	JP
FUJISAWA PHARMACEUTICAL	497	JP	SEPRACOR	243	US
GALEN	19	EU	SHIONOGI	249	JP
GEDEON RICHTER	35	EU	SHIRE PHARMACEUTICALS	201	EU
GLAXOSMITHKLINE	4,360	EU	SKYEPHARMA	46	EU
GUERBET	21	EU	STADA ARZNEIMITTEL	15	EU
IPSEN	122	EU	TAISHO PHARMACEUTICAL	235	JP
IVAX	1	US	TAKEDA CHEMICAL	990	JP
JOHNSON & JOHNSON	3,956	US	TANABE SEIYAKU	186	JP
KISSEI PHARMACEUTICAL	108	JP	TEVA PHARMACEUTICALS	163	IL
KYOWA HAKKO KOGYO	233	JP	UCB	246	EU
LUNDBECK	190	EU	WATSON PHARMACEUTICALS	81	US
MERCK CO	2,677	US	WYETH	2,080	US
MERCK KGAA	572	EU	YAMANOUCHI PHARMACEUTICAL	533	JP
MITSUBISHI PHARMA	401	JP	ZAMBON	27	EU

Chapter 3

Firm entry into new scientific areas:
The role of collaboration with insiders and outsiders[21]

ABSTRACT

As science is continuously evolving, the sustainability of firms' position in high-tech industries rests on their ability to recognize and act upon opportunities in new research areas. However, the abundancy and incompleteness of disclosed scientific output makes this a challenging task for firms to take-on autonomously. I posit that the social connections that firms can establish in scientific areas may help them to gather and process relevant information on new areas. The study investigates the role of different social connections to an area's community on firms' likelihood to enter the particular area. The empirical analysis examines the entry behavior of 133 biopharmaceutical companies within 290 disease areas between 1995 and 2002. The results signal that outsiders, scientists who do not publish in the area themselves, are inferior access points for information due to a lack of affinity, credibility, and information flow. In contrast, the information gathered from insiders -scientists actively publishing within the focal area - appear to strongly influence firms' probability on entry. While this information can directly originate from the insider, it can also reach the firm indirectly originating from the insider's co-authors in the area. Moreover, the results signal that insiders are effective brokers of their co-authors' information arguably thanks to their affinity with the area and the peer-review character of the scientific community. Hence, the study suggests that collaboration with highly-connected insiders is an effective strategy for firms' aiming to pursue opportunities within a new research area.

[21] This chapter is based on a single authored working paper.

3.1 INTRODUCTION

Scientific capabilities are considered key for sustained competitive advantage, especially within high-tech industries, like pharmaceuticals (Deeds, Decarolis, and Coombs 2000; Cassiman, Veugelers, and Zuniga 2008; Gambardella 1992). By performing research in-house, firms develop scientific capabilities and skills helping them to better absorb external information and knowledge. Additionally, they contribute to scientific understanding on which innovations can build. According to the resource-based-view of the firm, both this absorptive capacity and scientific understanding are strong competitive advantages as their tacitness, complexity and embeddedness make them rare, difficult to trade and, especially, inimitable to rivals (Deeds, Decarolis, and Coombs 2000; Fabrizio 2009).

A firm, subject to financial and organizational constraints, will focus on a subset of research areas in which it can develop these strong scientific capabilities (Ahuja, 2000; Mckelvey and Rake, 2016). However, as science evolves, opportunities may arise in areas outside the activity span of the firm. The sustainability of firms' competitive advantage rests on their ability to recognize and act upon these opportunities (Deeds, Decarolis, and Coombs 2000; Siciliano, Welch, and Feeney 2018; Iain M Cockburn et al. 2000). Though, the abundancy and incompleteness of disclosed scientific output (Stuart, Ozdemir, and Ding 2007; Cassiman, Veugelers, and Zuniga 2008) makes it hard for firms' to keep abreast with all scientific advances. Moreover, firms will only undertake the pursued of opportunities into new research after careful consideration as it is likely to entail the commitment of substantial resources for a long period (Song et al. 2018). Hence, when considering scientific capabilities as a key competitive advantage, it is important to understand how firms can become better informed on valuable research opportunities.

Scholars have theorized that collaboration with academia may be of strategic benefit for firms to remain informed on the latest scientific developments (Zucker and Darby, 1996; Rappert, Webster and Charles, 1999; Fabrizio, 2009; Hess and Rothaermel, 2012). While scientific output is publicly available, scientists may inform the firm on unpublished trial-and-error procedures, work in progress and early stage ideas (Fabrizio 2009; Agrawal 2006; Fuller and Rothaermel 2012; Sorenson, Rivkin, and Fleming 2006; Thursby and Thursby 2001). Moreover, the social network firms establish within the scientific community may not only act as an information gathering device but also as an information screening device (Ahuja 2000a; Belderbos, Gilsing, and Suzuki 2016b). While some studies in the alliance literature have indicated that collaboration may enhance firm's patenting rate in technological fields new-to-

the-firm (e.g. Gilsing *et al*. 2008; Guan & Liu 2016), quantitative studies on the role of scientific collaboration on firms' tendency to enter research areas new-to-the-firm is still lacking.

This study aims to verify the informative and guiding role of scientific collaboration within the biopharmaceutical industry. More precisely, I analyze how firms' likelihood to enter a new research area may depend upon its connectivity, established by scientific collaboration, to scientists active in the area. While I cannot identify whether increased connectivity may trigger firms interest for the area or whether connectivity strengthens firms initial intention on entry, I do argue that connections to scientists of the area provides firms access to information[22] valuable within the meticulous decision process.

I will distinguish between two groups of connections to scientists of the area: (1) insider connections where the information is transferred by scientists active in the research area and crosses the area boundary in its first-line communication to the firm, and (2) outsider connections where the information passes through a collaboration partner of the firm which is not active in the focal area and crosses the area boundary in the communication between the firms' collaboration partner and his contact in the area, i.e. the second line of communication. I posit that firms who are more strongly connected to the scientists of the new research area by insider connections are more likely to enter the research area than those who enhance their connectivity by outsider connection. First, along the lines of the broker[23] literature (Belderbos *et al.* 2016; Fleming *et al.* 2007; Quintane & Carnabuci 2016; Burt 2000; Salman & Saives 2005), I posit that the informative value of social connections is highly dependent on the senders' ability to process the information (absorb, evaluate, translate). Thanks to their own involvement, insiders have higher affinity to the area and are better known with its language, practices, and value criteria. Moreover, the flow of area relevant information to and from an outsider may be limited due to the high probability on frictions like lack of trust and credibility

[22] This study applies a strict definition of knowledge being tacit and embedded within individuals. In contrast, information is broadly defined as all communication that can be codified, e.g. ideas, unpublished findings, working papers, contact information.

[23] In this study the term broker refers to the definition of Gould and Fernandez (1989): someone who facilitates flows between two principals whether or not he attempts to extract a direct reward. While Gould and Fernandez (1989) specify five different broker roles, this study only reflects on two as it only investigates flows between groups. The first role examined is representative of the focal research area which is taken up by insiders because they belong to the same group as the sender of information or indirect tie. Second, the role of gatekeeper is taken up by the outsiders as they are just like the firm inactive in the research area and, hence, do not belong to the same group as the sender or indirect tie.

(Ghosh and Rosenkopf 2015; Reinholt, Pedersen, and Foss 2011; Perry-Smith and Mannucci 2016; Aral and Alstyne 2011).

In contrast to outsider connections, which by definition consist out of indirectly accessed scientists of the new area, insider connections include both area's scientists who collaborate directly with the firm and area's scientists who are indirectly accessed by other area members who collaborate with the firm. In other words, firms' collaborating with an area's scientist obtain a direct tie to that scientist and possibly multiple indirect ties to others active in the area. Direct ties have received a lot of attention in the university-industry literature (L G Zucker and Darby 1996; Lynne G. Zucker, Darby, and Armstrong 1998; Fabrizio 2009; Bruneel, D'Este, and Salter 2010). Close interaction facilitates the transfer of scientific information, which is highly complex and enables the transfer of tacit and embedded knowledge. Moreover, thanks to the trust established through direct interaction, information, especially on novel ideas and perspectives, is more likely to be shared between partners.

Although indirect ties have been ignored by most studies investigating university-industry collaborations, they are regarded of informative value within the social network literature (M. S. Granovetter 1973; Ahuja 2000a; Burt 2004). Indirect ties entail relatively lower maintenance costs as the firm can access multiple scientists indirectly while only investing in the collaborative relationship with one scientific partner. Burt (1992) even suggested that, opposed to establishing additional direct ties, the formation of many indirect ties may be an efficient strategy to be broadly informed. I propose that this reasoning only holds when the information of the indirectly connected scientist can sufficiently benefit from the background knowledge and trust between the direct tie and the firm. If not, the firm would be better informed by establishing additional direct ties. Within the context of this study, I state that the benefits of direct interaction can be extended to indirectly accessed information communicated to the firm. I namely focus on the information communicated by an insider who has overlapping complementary knowledge with her contacts in the same area (insider connections). Moreover, the scientific community has a culture that stimulates information exchange even without direct interaction.

This study examines entries into new scientific areas by analyzing the publications of the 133 largest R&D spenders in the biopharmaceutical industry from 1995 to 2002. Additional to the firm publications, we have information on all publications in Pubmed with corresponding research areas (MeSH terms) and disambiguated author names (Authority dataset: Torvik & Smalheiser 2009). Entry is determined by the first publication of the firm within a disease area

in which they were not active in the past 4 years. By investigating 290 disease areas (MeSH terms level 2 of branch C), 207,207 observations are taken into account with 3,169 registered entries. Scientists' activity in an area is determined by publications in the areas during the past four years. Social connections are established by having at least one co-publication within the four-year time-window. As scientists and firms publish in multiple research areas, firms can obtain a direct tie with an active contributor of a particular area by co-publishing with this scientist within a different area. The results indicate that increasing the share of an area's scientists that firms can connect to via outsiders not only has a smaller influence on firms' probability to enter the research area than increased connections through insiders but even has no significant impact on the probability of entry. Moreover, the results show that the number of indirectly connected insiders per directly connected insider has a positive relationship with firms' likelihood to enter the area.

This study contributes to the literature in several ways. First, this study analyzes the social network at the firm-area level providing some benefits in respect to the firm level analyses of prior work. I am able to distinguish between information that is communicated to the firm by insiders and outsiders to the research area. By also considering these knowledge boundaries, I acknowledge the diverging patterns found in knowledge versus collaboration networks (Guan and Liu 2016; C. Wang et al. 2014). By analyzing the issue at the level of the research area, the regressions will not explain the explorative nature of the firm as done by past research but will examine entry behavior into a specific area considering firm characteristics and area attractiveness. Moreover, the results emphasize the importance of the broker which has only received limited attention in the literature. Third, I investigate the importance of indirect ties within the uninvestigated context of entry into new scientific research areas. Ahuja (2000) has emphasized the importance of analyzing direct and indirect ties within different contexts. The results signal that the formation of indirect ties does form an effective strategy to remain informed on a distant scientific research area at least when the intermediary is an insider of the area. Finally, the study provides a first indication that, besides human capital, the social capital of academic collaboration partners, represented by their contacts, is valuable to firms in their quest for sustained competitive advantage. Generalized, the study shows that the behavior of actors can be related to the formation of indirect ties contradicting the devaluation of these ties by the study of Burt (2007) on second-hand brokerage and expanding the more nuanced view of conditional value mentioned by Shah and colleagues (2018).

3.2 BACKGROUND THEORY

3.2.1 The social network

The social network literature considers social connections between actors as pipelines of information and knowledge (Owen-Smith and Powell 2004c; Gilsing, Nooteboom, Vanhaverbeke, Duysters, and van den Oord 2008). In the early years of the literature stream, scholars were mainly focusing on the importance of close versus distant social connections for actors' performance. While some studies signaled the need for closure, others indicated the key importance of bridging structural wholes. Hence, scholars concluded that closure/density and distance/structural holes both have their advantages and disadvantages depending on the context and goal under investigation (Ahuja 2000a; C. Phelps, Heidl, and Wadhwa 2012; Burt and Merluzzi 2016). Density and closeness between actors enables trust, power and mutual understanding facilitating knowledge transfer and learning, while structural holes provide access to diverging opinions and new information facilitating recombination and knowledge creation (C. Phelps, Heidl, and Wadhwa 2012).

Recent studies have increased our understanding by relaxing the assumptions made in the pioneering work. First, while social connections provide opportunities for accessing information, this does not mean that the information is transferred and acted upon (Burt 2007). Consequently, scholars are investigating numerous elements that may influence the realization of effective knowledge transfer: i.e. repeated interaction generating trust (Aral and Alstyne 2011; J. Wang 2016a), social pressure stimulating cooperation (Shah, Levin, and Cross 2018) or knowledge overlap facilitating information absorption (Guan and Liu 2016). Second, accessing distant or previously unconnected alters does not necessarily imply access to new information. Similarly, close alters may embed new information. Studies that have compared the knowledge and social network of actors, have shown large structural differences between them and diverging benefits of certain ties and positions in each network (C. Wang et al. 2014; Guan and Liu 2016). To deal with this complexity, Shah, Levin and Cross (2018) have limited their view to providers of new information: social connections that cross an intra-organizational boundary.

In this paper, I will follow this strategy by only analyzing social connections to scientists active in a new research area. This implies that each social connection under investigation crosses the boundary of research areas or, in other words, they connect the community not active in the focal area, to which also the firm belongs, to the community active in the area. I will take a new perspective on these connections by identifying three different ways in which a firm may realize

such a boundary spanning connection. Moreover, I analyze these social connections within the novel context of university-industry collaborations. With the notable exception of *Belderbos et al.* (2016b), social network methods and concepts have received little attention in the university-industry literature. As explained in the following section, this study is especially interested in the informative role that academic collaboration partners may play in identifying or, after interest is established, evaluating research opportunities in new areas.

3.2.2 The informative role of academic collaboration partners

Despite the lack of hard empirical evidence, scholars have suggested that connections to scientists are beneficial for keeping abreast of scientific developments and estimating its value for the firm (Fabrizio 2009; Lynne G Zucker and Darby 1996; A. Hess and Rothaermel 2012; Rappert, Webster, and Charles 1999). Aggregated firm level studies have indicated that firms who perform scientific research in-house are able to build more rapidly on new scientific development and increase the quality of their patented innovations (Fabrizio 2009). Moreover, they have shown that the impact of in-house scientific research on firms' innovation performance is even greater when performed in collaboration with academia (Belderbos, Carree, and Lokshin 2004; Furman and MacGarvie 2009). Collaboration partners supposedly provide access to valuable information that, although existing in the firm's environment, would remain undetected by the firm: embedded skills and understanding, unpublished trial-and-error procedures and, early stage ideas and work in progress (Fabrizio 2009; Agrawal 2006; Fuller and Rothaermel 2012; Sorenson, Rivkin, and Fleming 2006; Thursby and Thursby 2001; Lynne G. Zucker, Darby, and Armstrong 1998). Zucker and Darby (1996) even described academic partners as "subcontractors to detect significant changes in basic science that might prove relevant to the firm".

While empirical studies verifying the belief in the guiding role of academic collaborators are still lacking, studies in other domains have signaled similar guiding roles of social connections. Research on strategic decision making has shown that information obtained through social connections is assigned greater weight by executives than information obtained from impersonal sources like written reports or scientific articles (Mcdonald and Westphal 2003). Past research on alliances has indicated that collaboration may enhance firm's patenting rate in technological fields new-to-the-firm (e.g. Gilsing *et al.* 2008; Guan & Liu 2016) and social ties between CEO's may influence strategic decision making (Mcdonald and Westphal 2003). Moreover, the study of Tzabbar (2009) has indicated that the recruitment of technologically distant scientists is related to significant technological repositioning of companies. However, it

can be questioned whether these findings can be translated to the setting of university-industry collaborations.

On the one hand, the firm will be confronted with the well-known difficulties of knowledge transfer within university-industry collaborations (Pe*rkmann et al.*, 2013b). In contrast to corporate partners and internal employees, who like the firm strive for commercial success and financial return, scientists are mainly driven by solving complex puzzles, obtaining social recognition and contributing to society (ribbon gold). These diverging research goals of firms and universities may influence the perceived risk on opportunistic behavior and trust between partners limiting successful knowledge transfer. Moreover, the culture of universities is significantly different from that of firms, so that information transfer is hindered by different languages and social norms (Belderbos, Gilsing, and Suzuki 2016a) (similar culture). On the other hand, the idea of homophily, often discussed in the social network literature, would indicate that firms are inclined to follow advice of more similar partners, like other firms or internal employees, than academic scientists. For instance, the study of Mcdonald and Westphal (2003) has indicated that CEOs are more likely to follow the opinion of like-minded contacts even when resulting in rationally inferior decisions. Because of cultural and motivational differences, scientists may be regarded unable to consider firms' needs and goals. Consequently, firms may be less inclined to value their advice and, even, ask for their advice in the first place.

3.2.3 Direct and indirect connections

Considering the cultural, institutional and cognitive differences between universities and industry, scholars have emphasized the benefit of direct interaction with academia for transferring knowledge and improving firms' innovation performance. Only by closely interacting with scientists can their highly complex, tacit and embedded knowledge be transferred (Zucker and Darby, 1996; Zucker, Darby and Armstrong, 1998; Fabrizio, 2009). Moreover, to cross the sectorial border between industry and academia, close interaction is of additional importance for the development of mutual understanding and trust (Bruneel, D'Este, and Salter 2010). In their study of the industrialization of biotechnology, Zucker and colleagues (1998) confirmed this key importance of direct ties by showing that firms' entry in biotechnology was triggered by collaboration with academic scientists in contrast to mere geographical proximity. Yet, as most studies on university-industry collaboration, this study did not investigate the influence of indirect ties.

Murray (2004), however, has shown in her case studies of entrepreneurial companies that academics' connections to other scientists, may also form an important resource for firms. Thanks to the academic partner, the firm is indirectly tied to her contacts, extending its access to scientific information. Also, in the social network literature benefits of indirect ties have been recognized (M. S. Granovetter 1973; Ahuja 2000a; Guan and Liu 2016). First, indirect ties increase the diversity and breadth of information accessed. Direct ties are said to transfer confirming opinions and similar knowledge. Indirect ties, however, hold new information and different perspectives by bridging to other social groupings. Thanks to indirect ties, firms may look at their capabilities and resources with a different lens helping them to see new opportunities (Guan and Liu 2016; Corey Phelps, Heidl, and Wadhwa 2012). Second, indirect ties may form an effective way to increase network size (Burt 1992). While each direct tie requires substantial investments in maintaining a good and trustworthy relationship, firms can benefit from many indirect ties while only covering the maintenance costs of one single direct tie. Third, firms may benefit from the intermediary or broker of an indirect tie (Belderbos, Gilsing, and Suzuki 2016b; Fleming, Mingo, and Chen 2007; Quintane and Carnabuci 2016; Ahuja 2000a; Burt 2000; Salman and Saives 2005). The intermediary will process the information he gathers: interpreting, evaluating and translating it (Song et al. 2018). The latter indicates that the information may be transformed by the intermediary in a more understandable language for the firm. Additionally, good brokerage can reduce processing costs and the risk on information overload at the firm as the intermediary will only pass through information in which he sees valuable opportunities for the firm.

Considering the benefits of indirect and direct ties, scholars have questioned whether they should be regarded as complements or substitutes. Burt (1992) suggested that many indirect ties might be regarded as an alternative strategy for a direct tie. In line with this reasoning, Ahuja's (2000) empirical analysis of the relationship between firm's alliance network and its innovation performance showed a substantially smaller marginal effect of indirect ties vis-a-vis direct ties, and a significant, though small, substitution effect between the number of direct and indirect ties. Other studies have, however, found diverging results (Corey Phelps et al. 2012). In their analysis of inter-organizational co-publication networks and the generation of new molecular entities in the treatment of cancer, Mckelvey and Rake (2016) found that indirect ties were of the highest importance. Then again, Gilsing and colleagues (2008), who analyzed inter-firm alliance networks and success in technological exploration, have emphasized the need for density indicating complementarity between direct and indirect ties. Within the same context,

Guan & Liu (2016) confirmed that both direct and indirect ties are related to more explorative patents.

These diverging results may be explained by the remarks of Ahuja (2000). First, he emphasizes that the benefit of direct and indirect ties is highly dependent on the analyzed context: the relationship between nodes, the content transferred and the examined goal. Second, he states that direct and indirect ties differ substantially in the nature or content of benefits that they provide to the focal actor. While both may transfer information, direct ties may also entail resource sharing (Ahuja, 2000) and knowledge transfer (Aral and Alstyne 2011). In short, the benefits of direct and indirect ties are not the same and should not be expected to react similarly within different contexts. Consequently, studies on the topic remain relevant to provide a better insight on their particular benefits and reactions. This study examines them within the uninvestigated context of entry into new scientific research areas and takes a novel firm-area perspective on the issue. In the following section, I will explain the setup of this study and the identification of different social connections. Following, I will discuss and formulate the hypotheses.

3.3 CONNECTIONS BY INSIDERS AND OUTSIDERS

While past studies on direct and indirect ties have proved their value, most of them have overlooked the fact that indirect ties can only be established through the formation of direct ties. Moreover, the broker literature has indicated that intermediates play an important role in information transfer *(Belderbos et al*. 2016a; Fleming *et al.* 2007; Quintane & Carnabuci 2016; Burt 2000; Salman & Saives 2005). In this study, we will consider this by identifying three different connections that a firm can establish to scientists active in a scientific area.

Figure 1 illustrates the social network of direct and indirect ties of a firm. Each actor presented in the figure (the firm and the eight scientists) have their own portfolio of research areas in which they actively publish, listed by the capital letters. Lines connecting these capital letters between actors represent some collaboration in the respective research area. Because actors publish in multiple research areas, they can collaborate in one area while sharing information on another area. In the figure, research area A divides all actors into two groups: (1) those actively publishing in the research area, and (2) those not actively publishing in the research area represented in grey. The firm is connected to seven of the eight scientists active in the area A. Five of these scientists are connected to the firm by insider connections of which scientists 5 and 6 directly and scientists 3, 6 and 8 indirectly. Two other scientists active in area A are

only connected to the firm by outsiders of the area. The firm can only obtain information of scientist 2 from scientist 1 and of scientist 10 from scientist 9.

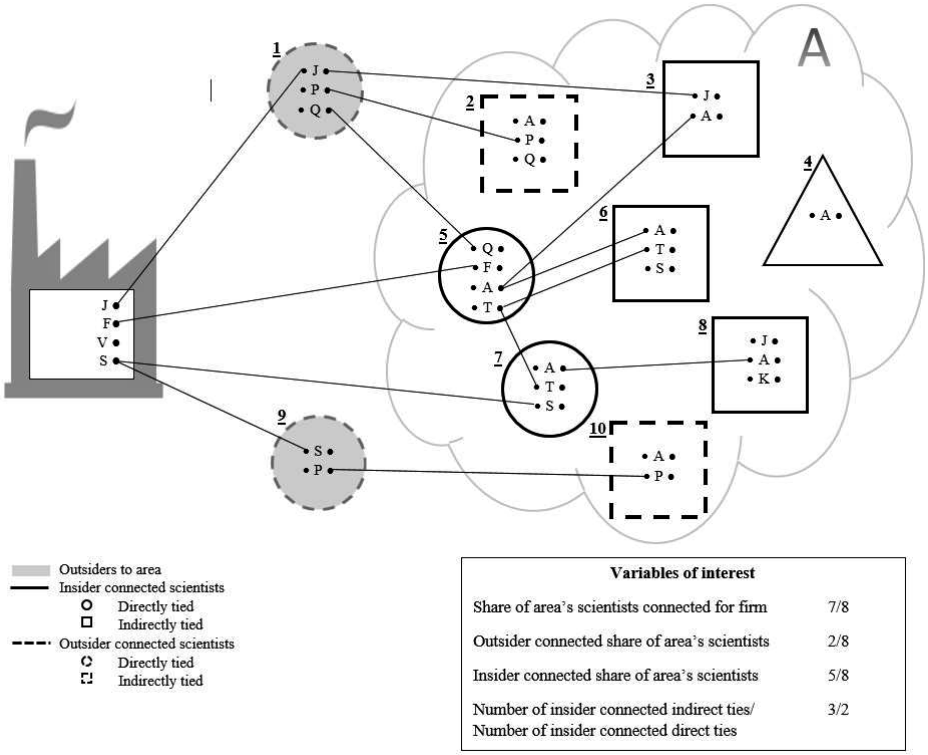


Figure 3.1. Firm's social network ties to a new scientific research area "A"

Prior research has applied firm-level analyses and has differentiated between direct (circles: scientists 1, 5, 7 and 9) and indirect ties (squares: 2, 3, 6, 8, 10) regarding all indirect ties as similar. By applying firm-area level analyses, I distinguish three different types of connections to scientists in the research area: direct connections of insiders (scientists 5 and 6), indirect connections of insiders (scientists 3, 6 and 8) and indirect connections of outsiders (scientists 2 and 10). In contrast to the information gathered by insiders, information of indirect connections of outsiders is communicated to the firm by a scientist who is not active in the research area. I will argue that the latter is an important difference to make when analyzing social connections. Within the following, I will build up the three hypotheses of this study based on the total number of connections to the research area, the difference between connections from insiders and outsiders and the interplay between direct and indirect ties among the connections of insiders.

3.4 HYPOTHESES

3.4.1 Total number of connections

As social connections are regarded pipelines of information and all three connections under investigation connect the firm to scientists who are active in the research area, it can be expected that firms with more of these connections will gather more information on the area. Zucker and Darby (1996) have suggested that information is crucial for firms entering a new scientific arena as it determines their ability to obtain, evaluate and use research opportunities. I posit that information may increase firms' likelihood to enter a research area in two ways. First, a higher quantity and coverage of information may increase the likelihood to find some novel perspectives or breakthrough ideas that can result in a valuable opportunity for the firm to combine, repurpose or generate scientific capabilities. Note that the search for opportunities is highly uncertain, limiting the possibility of focusing and planning the search process (Gilsing *et al*., 2008). While research results are disclosed within the scientific community, connections may provide access to information that is not yet published: ideas, trial-and-error procedures, or working papers. Moreover, past research has indicated that CEO's are more strongly influenced by personal advice than impersonal information, like reports, in their strategic decision making (Mcdonald and Westphal 2003). Second, social connections may provide general information of the research area to the firm: the distribution of expertise across scientists, common practices and background insights. Thanks to this meta-information, the firm may obtain sufficient confidence to grab the opportunity and take the risky leap of entry into an unfamiliar research area. Hence, I expect that by connecting to several members of an area's active community, the gathered information is more comprehensive increasing firms' likelihood to enter.

Hypothesis 1*: The probability that a firm enters a new scientific research area is greater, the higher the share of active scientists in the area who are member of the social network of the firm (directly or indirectly)*

3.4.2 Insider and outsider connections

The extent to which the accessed information will influence entry may rest on both the broker capabilities of those communicating the information to the firm and the severity of transfer frictions limiting the flow of information from one actor to another. A good broker is someone who can absorb, evaluate and translate information across disconnected nodes, like the firm and the new research area in the context of this study (Belderbos *et al*. 2016; Fleming *et al.* 2007; Quintane & Carnabuci 2016; Burt 2000; Salman & Saives 2005). The highly influential study

of Cohen and Levinthal (1990) states that people can absorb ideas more easily on topics matching their expertise. Thanks to overlapping knowledge, area-specific language is understood and holes in the transferred information can be filled (Lynne G Zucker, Darby, and Armstrong 2002; Fabrizio 2009). Moreover, to evaluate the quality and relevance of information, familiarity to the research area may also be of importance. Only when knowing the relevant background of ideas and research results, an evaluation on innovativeness and effectiveness can be made. For the broker task of translation, it is less clear whether the broker should have stronger cognitive proximity to the sender or the receiver of the information as translation entails understanding of both languages (Belderbos, Gilsing, and Suzuki 2016b). The investigation of all broker tasks signals that insider connections may have a stronger impact on firms' probability to enter a research area as they possess relevant knowledge to absorb and evaluate gathered information.

Furthermore, outsiders to an area may be more sensitive to transfer frictions, which reduces information flow (Ghosh and Rosenkopf 2015; Reinholt, Pedersen, and Foss 2011; Perry-Smith and Mannucci 2016; Aral and Alstyne 2011). As the scientist is not active in the research area, the focal scientist's contacts may assume that she is not interested and they may not be motivated to pass through their information. In a similar manner, the firm may not recognize the outsider as a possible source of information. More so, when the firm would obtain relevant information from an outsider, it may not respond with entry as its correctness and evaluation would not be trusted (Mcdonald and Westphal 2003). In other words, the outsider would not be regarded a credible source of information as he lacks personal experience in the focal area (Daniel Tzabbar 2009).

While recognizing these restrictions of outsiders in communicating the information of their contacts in the focal area to the firm, I believe that outsider connections to scientists within the area can positively impact entry of the firm. The outsider can still provide incomplete information which the firm could complement or verify with information from other connections. Most importantly, the outsider can provide contact information and recommendations of the firm to its contacts and, also, of its contacts to the firm in line with the tertius uingens strategy (Quintane and Carnabuci 2016). Consequently, I expect that connecting to a larger share of an area's community by outsider connections will have a positive though smaller influence on firms' probability to enter a research area than when achieved by insider connections.

Hypothesis 2: *The probability that a firm enters a new scientific research area is greater when increasing the share of active scientists in the area which are insider connections of the firm than when increasing the share of active scientists in the area which are outsider connections of the firm.*

3.4.3 Interplay of indirect and direct ties among the insider connections

Insider connections to a research area consist out of both direct and indirect ties. As explained above some scholars have emphasized the importance of direct ties (Zucker and Darby, 1996; Zucker, Darby and Armstrong, 1998; Fabrizio, 2009; Bruneel, D'Este and Salter, 2010), while others have highlighted the role of indirect ties (M. Granovetter 1985b; Burt 1997; Burt 2004; Mckelvey and Rake 2016). The latter have stressed that a search strategy based on many indirect ties may be more effective than the establishment of multiple direct ties as they demand substantially less maintenance costs.

However, after finding only a limited substitution effect, Ahuja (2000) has argued that these two strategies are far from perfect substitutes. He highlights that direct ties provide access to resources beside information. Additionally, others have signaled that close ties can also transfer interdependent instructions and contextual information (Aral and Alstyne 2011) leading to the transfer of more detailed and less noisy information (Belderbos, Gilsing, and Suzuki 2016b). In line with this reasoning, it can even be stated that direct ties enable the transfer of knowledge instead of mere information (Lynne G. Zucker, Darby, and Armstrong 1998). Additional to this benefit of transferring complementary knowledge, the close and bidirectional communication of direct ties provides trust between the sender and receiver of information. While trust is considered a main driver of collaborative learning (Valentina Tartari, Perkmann, and Salter 2014; Plewa et al. 2013; Bruneel, D'Este, and Salter 2010), it is regarded of special importance for sharing creative ideas and novel perspectives (Perry-Smith and Mannucci 2016). Trust ensures that the idea will move beyond the creator's mind by diminishing concerns about having the idea criticized or rejected. Moreover, it may stimulate credibility and further development of the idea as feedback from trustworthy partners is more easily accepted and perceived as useful.

When acknowledging these additional benefits of direct ties, the debate on the substitution between direct and indirect ties actually questions whether these benefits can be extended to information that is mediated by the direct tie. If extension is substantial, firms can scale up the information access of their social network by obtaining more indirect ties per direct tie with limited loss of benefits. Moreover, they would profit from the effectiveness of mediation: pre-

evaluation of information, diminished information overload and lower relation maintenance costs. This situation would support the idea of effective enlargement of information coverage through indirect ties as noted by Burt (1992). However when the additional benefits of direct ties cannot be extended, increasing the number of indirect ties linked to a direct tie may diminish the informative value of the firm's social network. In the following, I evaluate whether there is, within the connections of insiders, the possibility of extending the benefits of direct ties to the information they mediate.

As indicated above, a direct tie who is an insider of a particular research area is expected to be an effective mediator of the area's information. Not only do I expect that she is able to absorb and interpret the received information with considerably less noise, but thanks to her background knowledge on the area, she may also complement the external information with contextual information and interdependent instructions. Hence, I expect that the benefit of transferring complementary knowledge can be extended to mediated information originating from indirect ties. Moreover, as the scientific community is characterized by an open and peer reviewed culture, ideas and perspectives, are likely to be passed through by intermediates, even when subject to critique. In contrast to alliance networks, information flow is less bounded by contractual restrictions, opportunistic behavior, organizational barriers or secrecy. Hence, I reason that, within insider connections, the benefits of direct ties can be extended to the information they gather from their contacts. This entails that increasing the average number of indirect ties of the inside brokered direct ties is expected to increase the probability on entry into the research area.

Hypothesis 3: *The probability that a firm enters a new scientific research area is greater when its insider connections are characterized by a higher average number of indirect ties per direct tie*

3.5 DATA

The hypotheses are analysed using a publication dataset of the 133 largest R&D spenders in the biopharmaceutical industry from 1995 to 2002[24]. The firms have headquarters in the United States, Europe or Japan and are the largest R&D spenders (in absolute terms) in the

[24] The data covers 146 biopharmaceutical companies though 13 of them never enter a new research area. Consequently, they fall out of the sample as firm dummies are included in the regressions. Hence, I only report on the data of the 133 firms within this working paper.

pharmaceutical industry.[25] This industry is particularly interesting as the innovation performance of pharmaceutical companies has been linked to their tendency to collaborate with academia (Belderbos, Carree, and Lokshin 2004; Furman and MacGarvie 2009). However, more detailed investigation is still required to understand how academic partners assist firms in their innovation efforts. Their influence on the search for new research opportunities may be of particular interest. Because of the high number of discoveries within life sciences, firms are confronted with substantial search costs (Stuart, Ozdemir, and Ding 2007). Moreover, within the debate on the productivity crisis in the pharmaceutical sector scholars have suggested that the long hanging fruit has been picked and firms' search for new viable treatments has become increasingly difficult (e.g. Rafols *et al.*, 2014).

Another benefit of investigating the biopharmaceutical industry is its high data availability as explained in the introduction of this dissertation. This study will utilize publication data to determine firms' research activity and social network. Even though university-industry interactions may also occur through other channels, prior research has validated co-publications as a reliable indicator of collaborative research (Cockburn and Henderson 1998; Fabrizio 2009) and social network ties (e.g. Mckelvey and Rake, 2016).

The publication data which is gathered from PubMed is complemented by author disambiguation data (Authority dataset: Torvik and Smalheiser, 2009) and medical subject headings (MeSH terms of NIH). The latter enables the classification of publications within research areas. More precisely, these medical subjects are assigned to publications to facilitate search within the PubMed library of medical research. Most important, the MeSH terms are classified within a tree structure enabling the identification of scientific research areas. I focus on the subjects covering diseases (tree branch C) and utilize the second level terms within the tree leading to 290 disease areas under investigation[26]. It is important to recognize that this is one possible categorization which correctness is open for debate as science and scientific areas are highly complex and interrelated. I opted for this use of MeSH terms as it is based on an independent and expert panel and retrospectively updatesed across time. The latter entails that early publications of a new research stream which at time of development would be categorized

[25] As reported in the 2004 EU Industrial R&D Investment Scoreboard. This ranking lists the top 500 corporate R&D investors based in Europe, and the top 500 companies based outside Europe (mainly in US and Japan), in 2003.

[26] As scientific terminology and research areas evolve over time, adjustments to the MeSH tree are made every year and MeSH subjects within publication are updated retrospectively. For this study, I utilized the MeSH tree of 2014.

under an overarching area, are reclassified in a more suiting scientific category when the independence and uniqueness of the area has been recognized. As some research areas are more complex or receive more attention, the 290 disease areas under investigation show large differences in size, growth and development. However, in the analysis control variables are included to capture these interarea differences.

Publications are assigned to firms through listed affiliations[27]. As affiliation strings in publication databases are not unified and publications refer to affiliations of subsidiaries or divisions of a parent firm, publication data needs to be collected disambiguated at the consolidated parent-firm level by searching for names of the parent firm as well as all their majority-owned subsidiaries. Colen *et al.* (2018) provides a detailed description of the applied consolidation procedure. As some sample firms were only created after 1995 (e.g. Novartis was formed by a merger of Ciba-Geigy and Sandoz in 1996) and due to some missing values for R&D expenses, I rely on an unbalanced panel dataset in the empirical analyses.

By utilizing the author disambiguation data of Torvik and Smalheiser (2009), I determine the number of active contributors to a research area. Active contribution signals that the author has at least one publication with a citation within the research area during the past four years. I only consider recent contributions as research areas may evolve over time and non-frequent or past contributors may not be following the latest developments in the area which is of crucial importance for my research topic. Social ties are determined by co-authorship on publications of the past four years. Firms can obtain a direct tie to active contributors of an area, even when not (yet) active in the area itself, by collaborating with the co-author on a research topic that is not part of the focal research area. Moreover, a tie is solely determined by the presence of co-authorship and not by the frequency of copublication. In the appendix, a supplementary analysis does indicated the robustness of my findings after control for average tie strength of the direct connections.

[27] As PubMed has incomplete affiliation data, the firms' publications have been gathered within the Web of Science and, sequentially, linked to their counterpart within the PubMed dataset enabling the use of complementary datasets like Authority and MeSH.

Table 3.1. Descriptive statistics of variables of interest across all observation, observations of no-entry and observations of entry.

		All observations				Observations of no-entry				Observations of entry			
		Mean	**S.D.**	**Min.**	**Max.**	**Mean**	**S.D.**	**Min.**	**Max.**	**Mean**	**S.D.**	**Min.**	**Max.**
	DEPENDENT VARIABLE												
1	Dummy(Entry in research area)	0.02	0.12	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00
	VARIABLES OF INTEREST												
2	Share of area's scientists connected to firm	0.011	0.02	0.00	0.80	0.011	0.02	0.00	0.80	0.026	0.04	0.00	0.37
	Presence of outsider connected scientists	0.74	0.44	0.00	1.00	0.74	0.44	0.00	1.00	0.99	0.08	0.00	1.00
3	Outsider connected share of area's scientists	0.008	0.02	0.00	0.45	0.007	0.02	0.00	0.45	0.014	0.02	0.00	0.22
	Presence of insider connected scientists	0.39	0.49	0.00	1.00	0.38	0.48	0.00	1.00	0.97	0.18	0.00	1.00
4	Insider connected share of area's scientists	0.003	0.01	0.00	0.80	0.003	0.01	0.00	0.80	0.012	0.02	0.00	0.25
	Presence of insider connected indirect tied scientists	0.38	0.49	0.00	1.00	0.38	0.48	0.00	1.00	0.97	0.18	0.00	1.00
5	No. of insider connected indirect ties/ No. of insider connected direct ties	3.62	6.29	0.00	158.00	3.49	6.17	0.00	158.00	11.97	8.23	0.00	85.33
	FIRM CONTROLS												
6	Log(R&D $_{\text{t-1, million \$}}$)	4.03	1.60	0.22	8.49	4.02	1.59	0.22	8.49	4.67	1.74	0.72	8.49
7	Scientific relatedness of innovations $_{\text{patented in t}}$	4.82	5.92	0.00	41.00	4.82	5.92	0.00	41.00	4.71	5.50	0.00	41.00
8	No. of completed commercialization projects $_{\text{in t-10 to t-1}}$	9.30	25.44	0.00	208.00	9.17	25.19	0.00	208.00	17.28	37.46	0.00	208.00
9	Share of successfully completed projects $_{\text{in t-10 to t-1}}$	0.03	0.13	0.00	1.00	0.03	0.13	0.00	1.00	0.04	0.14	0.00	1.00

Variables in grey are not included in the regressions.
All variables are significantly different when comparing observations of entry and no-entry, except "Scientific relatedness of innovations".

Table 3.1 continued.

10	Scientific diversity of firm	11.51	8.05	0.00	36.31	11.46	8.04	0.00	36.31	14.65	8.32	0.00	36.31
11	No. of publications within overarching research area $_{\text{in t-4 to t-1}}$ / R&D $_{\text{t-1}}$	0.04	0.15	0.00	5.25	0.04	0.15	0.00	5.25	0.06	0.19	0.00	3.90
12	Basic research focus of publications in overarching research area $_{\text{in t-4 to t-1}}$	0.31	0.42	0.00	1.00	0.31	0.42	0.00	1.00	0.43	0.42	0.00	1.00
	RESEARCH AREA CONTROLS												
13	Number of active contributors in area/1000	15.00	35.23	0.00	554.14	14.28	33.61	0.00	554.14	61.18	79.09	0.00	554.14
14	Competition in research area	0.51	0.39	0.00	0.97	0.50	0.39	0.00	0.97	0.84	0.20	0.00	0.97
15	Dummy(growing research area)	0.72	0.45	0.00	1.00	0.72	0.45	0.00	1.00	0.87	0.34	0.00	1.00
16	Breadth of research area	25.04	67.23	0.00	612.00	23.89	64.58	0.00	612.00	98.89	146.57	0.00	612.00
17	Age of research area	11.84	6.99	0.00	55.00	11.82	6.97	0.00	55.00	13.51	8.09	0.00	54.00
	Basic research focus of research area	0.54	0.14	0.00	1.00	0.54	0.14	0.00	1.00	0.53	0.08	0.15	1.00
18	Basic research focus of research area $_{\text{normalized}}$	-0.00	1.00	-3.89	3.37	0.00	1.00	-3.89	3.37	-0.07	0.60	-2.77	3.37
	FIRM-AREA CONTROLS												
	Distance of research area to firm	1.00	0.01	0.47	1.00	1.00	0.01	0.47	1.00	0.99	0.02	0.73	1.00
19	Distance of research area to firm $_{\text{minimal, normalized}}$	0.00	1.00	0.25	3.95	0.00	1.00	0.25	3.95	-0.12	0.72	0.25	3.95
20	Av. quality of insider connected direct ties	0.15	0.84	-1.09	36.77	0.14	0.82	-1.09	36.77	0.74	1.57	-0.73	24.03
	DUMMIES												
	Year dummies												
	Firm dummies												
	Overarching research area dummies												

Variables in grey are not included in the regressions.
All variables are significantly different when comparing observations of entry and no-entry, except “Scientific relatedness of innovations”.

3.6 VARIABLES AND DESCRIPTIVE STATISTICS

3.6.1 Dependent variable and regression analysis

Table 1 lists the dependent variable and variables of interest with their descriptive statistics across all observations, observations of no-entry and observations of entry. The dependent variable is a dummy indicating whether a particular firm entered the research area within a certain year. Entry is defined as the firm's first scientific contribution to the area by means of at least one publication after absence to the area (no publications) in the past four years. This study does not consider repeated entries meaning that firm-area observations are only included within the sample when the firm has no publications in the area since 1990 or has entered with a first publication in the focal year. Although only 0.0153 % of the observations represent an entry, there is no need for rare event analysis as the absolute number of entries is substantial (3,169) (Williams 2017). Nonetheless, I utilize Complementary Loglog (Cloglog) analyses which is better able to deal with low probability of success than Logit or Probit analyses. Moreover, as it can be derived from the baseline hazard function, it is utilized as a discrete-time duration model (Jenkins 2005). As entry is observed per research area for each firm inactive in the area up until entry or until the year 2002, a panel regression analysis is applied with random effects and clustered standard errors per area-firm. Moreover, beside the control variables listed below, firm dummies and overarching research area dummies will be included in the models to account for unobserved characteristics of the firms and the research areas which may explain entry behavior.

3.6.2 Variables of interest

The variables of interest are based upon firm's social connections to the scientists actively publishing in the focal scientific research area. The first variable of interest, Share of area connected to the firm, covers all co-authors (direct ties) or contacts of co-authors (indirect ties) that a firm may have within the research area. This total set comprises connected scientists which are only reached though an outsider (Outsider connected share of area's scientists) and those of which the information can reach the firm by an insider of the area (Insider connected share of area's scientists)[28]. All these variables are scaled by the size of the area resulting in

[28] As network ties can be redundant, an active contributor of a research area can fall in one, two or three of the identified types of connections. While it is unrealistic to think that a firm would gather information which originates from one of its direct ties through another of its collaboration partners, insider connected direct ties are assigned first. Second, I opted to give priority to insider connected indirect ties as I expect that a firm is more likely to demand for information on an area from its direct tie that is active in the area than from one of its direct ties which are not active in the area. To test the

shares of the area's active scientists. This scaling is important as there is high heterogeneity across research areas in the number of active scientists: mean of 15,000 and standard deviation of 35 230. Hence, having connections to a certain number of scientists is likely to have a highly different meaning across these areas. The fourth variable of interest (Number of insider connected indirect ties/ number of insider connected direct ties) takes the ratio of the number of indirect ties and the number of direct ties within the total set of insider connected scientists. Hence, this variable will always be zero when the firm has no co-author (direct tie) who is active in the area. Table 1 shows that most observations of entry do, not only, record all three types of social connections, but also, consistently report higher shares than the observations of no-entry.[29] Furthermore, for the average number of indirect ties per insider connected direct tie, observations of entry register significantly and considerably higher means.

Figure 2 shows the mean values of the variables of interest for each year up until entry is observed. The presented lines indicate that the share of area's scientists connected to the focal firm increases the closer the firm is to entering the area. Moreover, while overall firms appear to establish most connections through outsiders, they seem to establish connections through insiders at a stronger pass when approaching entry. Interestingly, while the total share of area's scientists connected through insiders increases steeply, the share connected to direct ties with insiders augments only slightly. Hence, the figure provides a first indication that insider connections may be more strongly related to entry than outsider connections and that this relationship may be mainly driven by indirect ties.

To have a more detailed view on the latter, Figure 3 plots the number of direct ties to scientists of the focal research area and the average number of indirectly tied scientist of the area per directly tied scientist. First, this figure more clearly visualizes the increase in directly connected scientists towards entry presented by the climbing solid line. Second, the dashed line shows that these scientists themselves have an increasing number of social connections within the area leading to a steep increase in the average number of indirectly tied insider connected scientists per directly tied insider. This figure suggests that, when building up to entry into a research area, firms are establishing more and more collaborations with increasingly better-connected

robustness of my results upon this last assumption, I performed a robustness check where priority is given to the outsider connected indirect ties which results in similar findings.

29 Note that the maximum values are consistently higher for the observations of no-entry indicating the presence of an outlier in this subsample. However, supplementary analyses have indicated robustness to the exclusion of outliers.

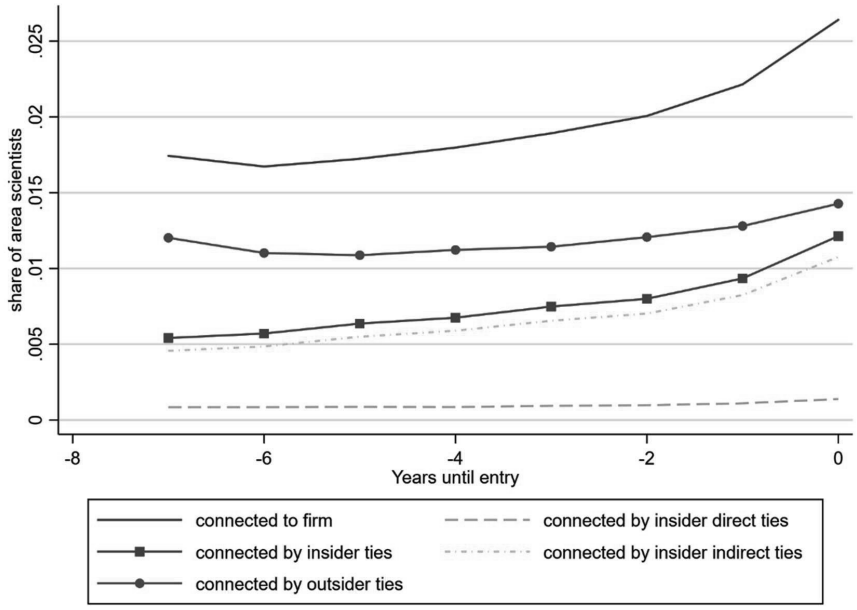


Figure 3.2. The formation of connections to the area across the years before entry

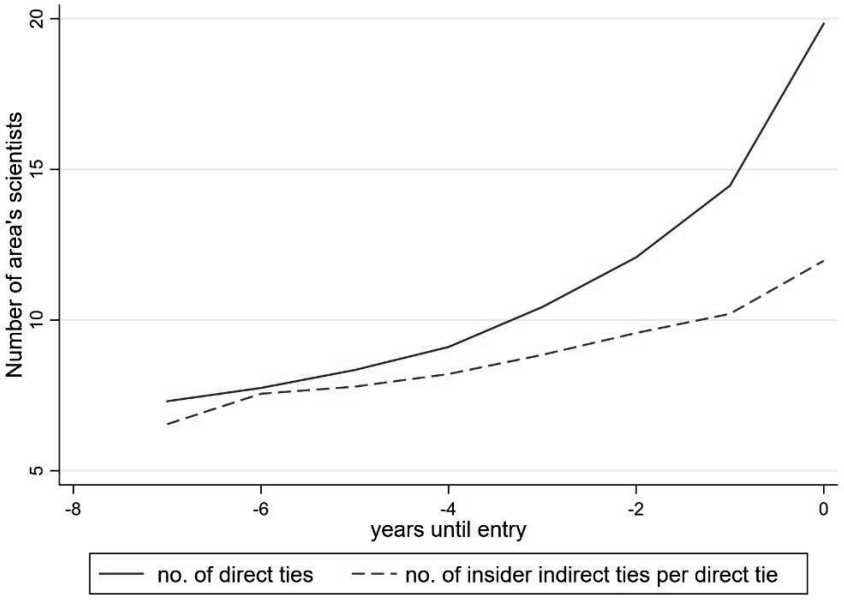


Figure 3.3. The formation of insider connections to area's scientists across the years before entry

scientists of the focal area. Also, Balachandran and Hernandez (2018) have also noted this creeping strategy among firm: low embedded firms form alliances with more embedded firm with the aim of reaching even more embedded actors.

3.6.3 Control variables

The controls listed in Table 1 consist out of firm level, area level and firm-area level variables. At the level of the firm, I control for differences in the size of firms' R&D investments, measured as one-year lagged R&D expenditures (expressed in million USD). Firms that marshal more R&D investments are more likely to explore (Leten, Belderbos, and Van Looy 2016).

The second control approximates the relatedness of the firm's innovations by taking the average number of scientific non-patent references on its patents applied for in year t. Firms of which the innovations are more strongly related to scientific research may find it more important to stay abreast with new scientific developments and more able to recognize and capture new scientific opportunities. However, this control is the only variable in Table 1 that does not significantly differ across the observed samples of no-entry and entry. This may imply that there is significant difference between using science, measured by non-patent references and performing science, as indicated by entry into new research areas.

The following two variables control whether firms may be more willing to enter a new research area when there pipeline is drying out due to many finalized commercialization projects or when they need to search for new opportunities after realizing many commercialization successes in their current markets. In the latter situation, firms may not only be able to take on more risky research projects within their portfolio, they may also want to avoid the exploration within their existing research area in fear of cannibalizing their newly launched products.

Scientific diversity is the following firm-level control variable and is based on the reverse of the Herfindahl index of firms' publication shares across the 290 disease areas in t-4 to t-1. A firm who is more diversified in its scientific efforts may be more experienced in exploration and may have developed a more divers knowledge base increasing the number of possible re-combinations that can be made with external knowledge. In contrast, high scientific diversity may also indicate greater potential for information overload limiting firms ability to recognize and capture new opportunities (March, Simon, and Guetzkow 1993; Kim *et al.* 2018).

The final two firm-level control variables are based upon firms' publications in the overarching research class of the focus area (first level of the MeSH tree). The number of publications within

this overarching research area relative to firms' R&D budget approximates the importance of related research within the past activities of the firm signaling interest in the broader topic. Moreover, the variable may indicate some "intellectual skills" relevant to the focal area as past research has emphasized that any piece of knowledge can only be acquired by people who possess certain prerequisite pieces, building from one level of knowledge to the next *(Song et al.* 2018). Finally, a control for firms' focus on basic research within this overarching research is included. A high focus on basic research shows that the firm's intentions are more likely to be based on discovery and exploration in contrast to commercialization, upscaling and exploitation. This variable is measured by the share of basic research (CHI journal categorization levels 3 and 4) within the firm's publication portfolio in the overarching research area of the past four years.

For the research area, it is important to control for reasons that may influence its attractiveness to firms searching for a new competitive advantage. As the size of a research area may signal its importance for society, the number of active contributors in the research area forms a first control. Moreover, the variables of interest are scaled by this measurement of size. Second, as indicated by studies on entry in technological fields, firms may be more reluctant to enter a research area whit a dominating player and high entry barriers (Leten, Belderbos, and Van Looy 2016; Polidoro 2013). Hence, we include a variable of competition in the research area calculated by 1 – the Herfindahl index of market shares of the active firms within the research area in the past four years. The third control variable at the level of the research area indicates that the yearly number of publications within the area has been growing during the past four years. Research areas which are enlarging may attract attention and may signal the presence of opportunities. Furthermore, I control for the breadth of the research area. Within the used taxonomy of research areas (MeSH level 2) some areas contain many subtopics (max. 612) while others are not further disentangled in subcategories. It is likely that broader areas may entail more opportunities both scientifically as commercially: multiple patient groups, treatment opportunities. Age is the fifth research-area control measured by the number of years from the first publication within the NUTS2 regions in which our sample firms are active and the year under examination. The last control is the basic research focus of the research area and is approximated by the number of basic research publications (CHI journal classification 3 and 4) divided by the number of applied research publications (CHI journal classification 1 and 2) within the area. This variable will be included in the regression analysis after normalization to assist in more natural interpretation of the coefficient. A research area that is already more

mature is expected to form an 'easier' subject for entry as knowledge is already more developed and codified.

Additional to these controls, I also include the cognitive distance between the research area and the past research of the firm. To accomplish this, I calculated a Jaccard Similarity Index between the 290 disease areas based on their co-occurrence matrix during the past four-year window. This implies that I allow areas to become less or more distant over the sample period of 1995 to 2002. Firms' cognitive distance to the focal area is then determined by the minimum similarity to the research areas it has been active in the past four years. It is well known that firms will be less likely to enter a research area which is more distant as it limits possibilities of complementary knowledge and synergy *(Zucker, Darby and Armstrong, 2002; Gilsing et al.*, 2008; Guan and Liu, 2016). This variable has been normalized as it has a very low variance within areas that are new to firms. As final control, I include the average quality of the firm's insider connected direct ties to the area. As scientists with a higher number of connections may also be of higher quality (Coleman 1988a; Hagedoorn and Wang 2012), this control is important for the internal validity of the variable that compares the number of indirect ties per direct tie. The research quality of a direct tie is measured as the total number of citations (until July 2009) received by her articles published in the research area during the past four years. As research areas may differ in their citation and publication tendency (S. Kelchtermans and Veugelers 2011), this quality has been normalized per year and disease area. Table 1 shows that all control variables are significantly higher for observations of entry, except distance and immaturity that are lower.

Additional to the control variables mentioned in the table, the analyses will also include several dummies. Time trends and overall chocks are covered by year dummies. Parent dummies control for firm differences in their strategic openness to the exploration of new research areas. And finally, dummies for the 26 overarching disease groups are included to shield the analyses from omitted technical characteristics like the subject (group of animal diseases) or the technicality of treatment (group of wounds and injuries). To conclude, the correlation matrix shown in Table 2 shows four correlations above 0.6. The first two occurrences refer to variables of interest that are not jointly incorporated in one regression model. The other two high correlations reflect on the inevitable linkage between breadth or diversity and size. As each of these variables has its own theoretical ground for inclusion and their individual relation to the

Table 3.2. Correlation matrix of the regression variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.000																		
2	**0.079**	1.000																	
3	**0.051**	**0.917**	1.000																
4	**0.094**	**0.827**	**0.534**	1.000															
5	**0.165**	**0.184**	**0.111**	**0.235**	1.000														
6	**0.050**	**0.533**	**0.564**	**0.335**	**0.113**	1.000													
7	-0.002	**-0.078**	**-0.083**	**-0.049**	**0.015**	**-0.288**	1.000												
8	**0.039**	**0.526**	**0.554**	**0.335**	**0.038**	**0.530**	**-0.141**	1.000											
9	**0.007**	**0.035**	**0.036**	**0.024**	**0.008**	**0.103**	**-0.111**	**0.036**	1.000										
10	**0.049**	**0.517**	**0.543**	**0.330**	**0.145**	**0.739**	**-0.174**	**0.427**	**0.081**	1.000									
11	**0.015**	**0.042**	**0.030**	**0.046**	**0.064**	**-0.126**	**0.088**	**-0.036**	**-0.024**	**0.068**	1.000								
12	**0.037**	**0.231**	**0.234**	**0.161**	**0.125**	**0.350**	**-0.084**	**0.170**	**0.036**	**0.427**	**0.260**	1.000							
13	**0.163**	**-0.089**	**-0.111**	**-0.032**	**0.452**	**-0.146**	**0.047**	**-0.081**	**-0.021**	**-0.160**	-0.002	**-0.066**	1.000						
14	**0.107**	**-0.042**	**-0.073**	**0.015**	**0.400**	**-0.130**	**0.041**	**-0.097**	**-0.012**	**-0.128**	**0.019**	**-0.035**	**0.397**	1.000					
15	**0.041**	-0.006	**-0.014**	**0.007**	**0.123**	**-0.040**	**0.014**	**-0.026**	-0.005	**-0.040**	**-0.007**	**-0.033**	**0.163**	**0.185**	1.000				
16	**0.137**	**-0.079**	**-0.096**	**-0.031**	**0.377**	**-0.124**	**0.038**	**-0.070**	**-0.019**	**-0.136**	**-0.013**	**-0.077**	**0.797**	**0.340**	**0.154**	1.000			
17	**0.030**	-0.004	**-0.009**	0.005	**0.142**	**-0.028**	**0.035**	**-0.011**	0.001	**-0.018**	0.004	**-0.012**	**0.095**	**0.330**	**0.097**	**0.132**	1.000		
18	**-0.009**	-0.003	**-0.014**	**0.014**	**-0.067**	**0.014**	-0.004	**0.015**	0.002	**0.013**	**0.026**	**0.067**	**-0.057**	**-0.164**	**-0.154**	**-0.058**	**-0.150**	1.000	
19	**-0.016**	**-0.037**	**-0.040**	**-0.021**	**-0.083**	0.006	**-0.033**	**-0.030**	**0.007**	**-0.015**	0.000	**0.019**	**-0.056**	**-0.163**	**-0.141**	**-0.052**	**-0.207**	**0.279**	1.000
20	**0.088**	**0.093**	**0.055**	**0.119**	**0.348**	**0.034**	**0.049**	**0.022**	-0.006	**0.044**	**0.027**	**0.037**	**0.105**	**0.106**	**0.028**	**0.096**	**0.040**	**-0.024**	**-0.024**

Grey shaded cells present correlations above 0.7.

dependent variable does not form the focus of this study all variables are maintained within the regression analysis.

3.7 REGRESSION RESULTS

The regression results are presented by Table 3 and 4 including coefficients, z-values and o-values. I apply a significance threshold of 0.05 for the p-values as shown by the coefficients in bold. As the results originate form Cloglog regression models, they needs to be interpreted with some caution. The marginal effect of a variable namely depends on the value of the other variables in the regression implying that a variable with a significant coefficient may still have marginal effects equal to zero, or even with an opposite sign. For the control variables, I will discuss the coefficients presented in tables 3 and 4. However, for the variables of interest it is useful to have a closer look at the marginal effects across observations within the sample. Table 6 presents the average marginal effects with their 0.95 confidence intervals. As the sample represents values observed in reality, this method provides a more realistic view on the importance of the variable than merely looking at the marginal effect at the mean.

Model 1 of Table 3 only contains the control variables. Even when including firm dummies within the regression as done within Model 2, R&D expenditures has a positive impact on firms' probability to enter a new research area. This may indicate that firms who increase their investments in research are looking for new ways to expand their business while firms of which the R&D department is shrinking are focusing on increasing efficiency by exploitation of existing knowledge. Moreover, firms which relative to their size are more active in research overarching the focal area and of which this activity is of a more basic nature, are more likely to enter the area. Moreover, the remaining firm-level controls are insignificant or turn insignificant after inclusion of the firm dummies in Model 2. The area-level controls are all highly significant. Entry appears more attractive in research areas that are large, highly competitive, growing, broader, older, and have a stronger basic research focus. These findings are in line with the formulated expectations. Overall firms appear most confident in the development of competitive scientific capabilities within the growth phase of the area's life cycle where no dominant competitor is yet present and fundamental understanding is still building up, though were uncertainty is diminishing signaled by increasing age, size and broadness. Finally, the cognitive distance between the firms' research and the area has no additional explanatory power within the model. As the variables on firms' research activity in the overarching research area were significantly positive, these results may indicate that vertical

Table 3.3. Cloglog panel regression with random effects predicting firm's probability to enter a research area

	Model 1		**Model 2**		**Model 3**	
	Coef.	(z) p	Coef.	(z) p	Coef.	(z) p
VARIABLES OF INTEREST						
Share of area's scientists connected to firm					10.48	(11.46) 0.00
FIRM CONTROLS						
Log(R&D $_{t-1, \text{million \$}}$)	0.34	(14.29) 0.00	0.29	(5.35) 0.00	0.29	(5.12) 0.00
Scientific relatedness of innovations	0.02	(6.34) 0.00	0.01	(2.39) 0.02	0.01	(2.37) 0.02
Number of completed commercialization projects	0.00	(6.45) 0.00	-0.00	(-1.18) 0.24	-0.01	(-3.05) 0.00
Share of successfully completed projects	0.33	(2.33) 0.02	0.09	(0.28) 0.78	0.07	(0.20) 0.84
Scientific diversity	0.06	(13.91) 0.00	-0.00	(-0.44) 0.66	-0.00	(-0.25) 0.81
No. of publications within overarching research area/R&D	0.64	(8.77) 0.00	0.50	(5.78) 0.00	0.48	(5.31) 0.00
Basic research focus of publications in overarching research area	0.23	(4.24) 0.00	0.17	(3.03) 0.00	0.19	(3.38) 0.00
RESEARCH AREA CONTROLS						
Number of active contributors in area	0.01	(13.66) 0.00	0.01	(14.26) 0.00	0.01	(14.36) 0.00
Competition in research area	3.87	(25.03) 0.00	3.92	(25.16) 0.00	3.97	(24.73) 0.00
Dummy(growing research area)	0.30	(5.43) 0.00	0.31	(5.55) 0.00	0.30	(5.36) 0.00
Basic research focus of research area	0.00	(3.06) 0.00	0.00	(3.10) 0.00	0.00	(2.79) 0.01
Breadth of research area	0.01	(3.31) 0.00	0.01	(3.29) 0.00	0.01	(2.95) 0.00
Age of research area	0.17	(4.90) 0.00	0.17	(4.82) 0.00	0.11	(3.04) 0.00

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.3 continued.

	Model 1		Model 2		Model 3	
	Coef.	(z) p	Coef.	(z) p	Coef.	(z) p
FIRM-AREA CONTROLS						
Distance of research area to firm	-0.04	(-1.45) 0.15	-0.04	(-1.45) 0.15	-0.04	(-1.25) 0.21
Average quality of insider connected direct ties	No		No		No	
DUMMIES						
Year dummies	Yes		Yes		Yes	
Firm dummies	No		Yes		Yes	
Overarching research area dummies	Yes		Yes		Yes	
Constant	-10.30	(-42.10) 0.00	-8.34	(-18.69) 0.00	-9.05	(-19.42) 0.00
N	207207		207207		207207	
LL	-13085.64		-12667.12		-12592.57	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

relatedness which provides general understanding to absorb more refined information, is more important for the entry decision than horizontal relatedness providing complementary information and recombination potential.

Within Model 3, the first hypothesis of this study is tested by inserting the share of scientists within the research area who are connected to the firm. The corresponding coefficient is significant and positive. The descriptive statistics on the marginal effect of this variable, shown in Table 5, indicate that the marginal effect is always larger than zero and has an average marginal effect of 0.160 within the sample.

In Model 4 of Table 4, the share of connected scientists within a research area is separated in those accessed through an insider and an outsider. Both the average marginal effects shown in Table 5 as the coefficients in Table 3 demonstrate that outsider connected scientists have a significantly negative influence on firms' entry rate. In contrast, insider connected scientists have a significant positive impact on firms' likelihood to enter the research area. Hence, the overall effect of connections that was found in Model 3 hides an important distinction based on the connection through which the area's scientists are reached. The findings do strongly support Hypothesis 2 stating that firms' probability to enter a research area is larger when the relative size of its social network within the area is increased by insider connections than when it is increased by outsider connections.

Table 3.4. Cloglog panel regression with random effects predicting firm's probability to enter a research area, examining the social connections to scientists in the area

	Model 4		**Model 5**		**Model 6**	
	Coef.	(z) p	Coef.	(z) p	Coef.	(z) p
VARIABLES OF INTEREST						
Outsider connected share of area's scientists	-5.72	(-2.31) 0.02	-4.86	(-2.06) 0.04	-4.53	(-1.91) 0.06
Insider connected share of area's scientists	15.38	(6.27) 0.00	14.41	(6.69) 0.00	14.65	(6.84) 0.00
No. of insider connected indirect ties/ no. of insider connected direct ties			0.04	(14.42) 0.00	0.03	(11.54) 0.00
FIRM CONTROLS						
Log(R&D $_{\text{t-1, million \$}}$)	0.29	(5.24) 0.00	0.26	(4.60) 0.00	0.27	(4.65) 0.00
Scientific relatedness of innovations	0.01	(2.38) 0.02	0.01	(1.80) 0.07	0.01	(1.85) 0.06
Number of completed commercialization projects	-0.00	(-1.33) 0.18	-0.00	(-1.31) 0.19	-0.00	(-1.37) 0.17
Share of successfully completed projects	0.08	(0.25) 0.80	0.16	(0.48) 0.63	0.14	(0.42) 0.68
Scientific diversity	-0.00	(-0.36) 0.72	-0.00	(-0.38) 0.70	-0.00	(-0.12) 0.91
No. of publications within overarching research area/R&D	0.49	(5.44) 0.00	0.47	(5.07) 0.00	0.47	(4.94) 0.00
Basic research focus of publications in overarching research area	0.18	(3.07) 0.00	0.16	(2.79) 0.01	0.16	(2.82) 0.00
RESEARCH AREA CONTROLS						
Number of active contributors in area	0.01	(13.72) 0.00	0.01	(12.26) 0.00	0.01	(12.78) 0.00
Competition in research area	4.06	(24.06) 0.00	3.80	(22.93) 0.00	3.86	(22.87) 0.00
Dummy(growing research area)	0.31	(5.45) 0.00	0.29	(5.07) 0.00	0.30	(5.22) 0.00
Basic research focus of research area	0.00	(2.81) 0.01	0.00	(2.25) 0.02	0.00	(2.04) 0.04
Breadth of research area	0.01	(3.27) 0.00	0.01	(3.27) 0.00	0.01	(3.33) 0.00
Age of research area	0.14	(3.66) 0.00	0.12	(3.25) 0.00	0.13	(3.33) 0.00

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.4 continued.

	Model 4		Model 5		Model 6	
	Coef.	(z) p	Coef.	(z) p	Coef.	(z) p
FIRM-AREA CONTROLS						
Distance of research area to firm	-0.05	(-1.61) 0.11	-0.05	(-1.47) 0.14	-0.05	(-1.54) 0.12
Average quality of insider connected direct ties	No		No		0.13	(6.47) 0.00
DUMMIES						
Year dummies	Yes		Yes		Yes	
Firm dummies	No		Yes		Yes	
Overarching research area dummies	Yes		Yes		Yes	
Constant	-8.73	(-18.67) 0.00	-8.57	(-18.09) 0.00	-8.80	(-18.28) 0.00
N	207207		207207		207207	
LL	-12547.81		-12397.67		-12347.81	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.5. Descriptive statistics on the marginal effects of the variables of interest considering the other variables as observed within the sample

Marginal effect (and 95% conf. interval) of the variable:	**Considering the logit predictions of:**			
	Model 3	**Model 4**	**Model 5**	**Model 6**
Share of area's scientists connected to firm	0.160 (0.131; 0.189)			
Outsider connected share of area's scientists		-0.877 (-0.163; -0.013)	-0.075 (-0.146; -0.004)	-0.070 (-0.141; 0.002)
Outsider connected share of area's scientists		0.236 (0.160; 0.313)	0.221 (0.154; 0.288)	0.226 (0.159; 0.291)
No. of insider connected indirect ties/ No. of insider connected direct ties			0.0006 (0.0006; 0.0007	0.0005 (0.0004;0.0006)

By including the variable 'Number of insider connected indirect ties/number of insider connected direct ties', the models 5 and 6 presented in Table 4 investigate the balance between indirect and direct ties among the area's scientists who are connected to the firm by an insider. As hypothesized, the focal variable has a significantly positive coefficient. Furthermore, the average marginal effect, presented in Table 5, shows that when comparing two similar research areas in which a firm is tied to the same share of its scientists by insider connections, the likelihood that the firm will enter is 0.0006 higher for the area in which the firm has an equal number of indirectly and directly tied scientists than for the area in which the firm has only directly tied scientists. In comparison to Model 5, Model 6 additionally controls for the average quality of the direct ties to scientists active in the research area. Inclusion of this control is important to settle concerns about correlation between human and social capital. While the outsider-connected share of area's scientists becomes insignificant, the results remain highly similar.

3.8 SUPPLEMENTARY ANALYSIS[30]

3.8.1 Performance after entry

In the above, I mainly discussed social connections as access points to information and argued that more informed firms are more likely to enter a research area as they can recognize research opportunities. However, social connections may also generate social pressure pushing firms to enter research areas at unfavorable times or in an unsuited manner. Alternatively, firms may obtain a publication in a certain research area without the intention to remain active. Hence, it is important to verify the findings by examining the relationship between firms' pre-entry social connections to a research area and firms' performance in this area after entry.

For a first supplementary analysis, I measure the persistency of firms' presence in the area by at least five publications in the eight years upon entry (t to t+7). Firms publish on average 3.74 articles in the area within the eight years after entry. Almost 25 percent of them have at least five publications during this period. The results of the Cloglog regression, presented in Table A1, are consistency with the studies main findings and show that insider connections to a larger share of an area's community does not only increase the likelihood on entering a research area but also on continued presence within that area. Hence, these results do not indicate that social connections place firms under pressure to enter research areas. In contrast, they appear to help

[30] The results tables of the supplementary analysis can be found within the appendix of this chapter.

firms in making a more informed entry decision with higher chances on sustained activity in the area.

Table 8 lists descriptive statistics of the publications observed in the first two years that a firm is active in a new research area. On average firms publish one or two publications in those first two years though we also observe a firm publishing 18 publications during this period. The majority of these publications (87%) is performed with an academic partner. Moreover 57 percent of the publications results from joint research with an old collaboration partners that was already active in the area (direct insider connections) and 2 percent is co-published with an area contact of old collaboration partners in the area (indirect insider connections). In contrast, it seems less likely that the firm starts a collaborative relationship with area contacts of old collaboration partners who were not active in the area (0.6%, outsider connections). Similar to analyzing the portfolio of publications, the composition of the co-authors exists for a substantial part out of previous direct insider connections. Finally, the last row in the column indicates that only in ten percent of the observed entries the firm performs all the research alone.

In a second set of complementary analyses, I examine these first-year publications in more detail by running two-stage regression analyses. Table A2 presents the results of second-stage Poisson regressions, where the Mills ratio originates from the main Cloglog regression of this study. The variables that form the exclusion restriction consist out of firm controls which approximate firms openness to entry though are not related to scientific performance (scientific relatedness of innovations, number of competed commercialization projects, share of successfully completed projects), and area controls which are barriers for entry into the area though do not influence scientific output (competition in research area, dummy (growing research area, age of research area). The results of the first model in Table A2 signal that the social connections obtained before entry are still related to post-entry publication performance with similar to their relation to entry a negative coefficient for outsider and a positive coefficient for insider connections. However, as shown by the second and third model, the social connections are only related to collaborative publications after entry and not significantly related to research published by the firm alone. Additional to these findings, it is remarkable that the third variable of interest, the average number of insider connected indirect ties per insider connected direct tie is insignificant in all model implying indifference between brokered or directly obtained information from scientists knowledgeable on the area. Similarly, Burt (2007) found in his study of second-hand brokerage that the investigation of direct ties may be

sufficient to explain actors performance. More precise, his results indicated that brokerage within the direct tie is a necessary condition to benefit from brokerage in the indirect tie.

3.8.2 Strength of social connections

As knowledge transfer is facilitated by trust and face-to-face interaction (Zucker and Darby, 1996; Zucker, Darby and Armstrong, 1998; Fabrizio, 2009), it can be argued that the strength of established connections plays an important role in the analysis of social networks (Aral and Alstyne 2011; J. Wang 2016b). Table 10 presents two supplementary analyses including controls for the strength, average number of co-publication in the past four years, between the firm and the scientists on the path to the connected scientists of the focal area. I only take into account the strength of these first-stage connections to reflect the firm's relationship to the communicator of the information. In Model 1 of Table A3, I simply include the tie strength for the outsider and insider connections. In Model 2, I implement two variables for the insider connections: strength of the direct insider connections and strength of the indirect insider connections.

Our results appear robust to this inclusion of tie strength. Interestingly, they also show that stronger outsider connections decreases the probability on entry, while stronger insider connections elevates the likelihood on exploration of a new research area. Considering that a closer relationship entails more knowledge transfer thanks to frequent interaction and the build-up of mutual understanding, these dual results seem surprising. However, higher tie strength can also be related to trust and credibility. Advice coming from a trusted and credible individual will have a larger impact on an actor's action strengthening the initial positive and negative findings of respectively insider and outsider connections. Finally, Model 2 of Table A3 additionally presents stronger coefficients for the strength of indirectly accessed insider connections than for those directly accessed. In other words, tie strength is of special importance to obtain influential information from an area's representatives.

3.8.3 Connecting to star scientists

Past research has indicated that scientist with extraordinary research performance obtain an influential position within their social network: steering the emergence of the new biotech companies (Kehoe and Tzabbar 2015; Daniel Tzabbar 2009; Lynne G. Zucker, Darby, and Armstrong 2002a). Thanks to their reputation, they are perceived experts within their area and, hence, form highly credible sources of information. Moreover, some scholars state that they can only realize such extraordinary performance thanks to their ability to recognize and grasp

new opportunities (A. Hess and Rothaermel 2012). In contrast to non-star scientists, who do the hard work of exploitation, star scientists are assumed to have high exploratory skills. This would imply that they might provide the needed information at the right time opening up opportunities for entry to their collaborating firms. While these argumentations would suggest that star scientists are beneficial access points to information on an area, star scientists may also have their disadvantages. Studies on stars within firms have indicated that their elevated social position may imply the NIH syndrome limiting the growth of others and their ideas (Daniel Tzabbar 2009; Kehoe and Tzabbar 2015; B. Groysberg, Polzer, and Elfenbein 2011). Consequently, star scientist may not be open to information from others and unlikely to transfer others' knowledge to their contacts. Moreover, communication with stars scientists may be difficult for actors unfamiliar with the stars' research area as they are unlikely to fully understand the high levels of specialization (Shah, Levin, and Cross 2018). Finally, the second chapter of this dissertation has mentioned issues of limited commitment and time availability when collaborating with academic star scientists.

To analyze the impact of stardom within our analyses, each variable of interest has been recalculated for star[31] and non-star linked connections. More precise, star-linked outsider connections are area's scientists who are contacts of a star scientist who is not active in the area herself. Star-linked insiders are star scientists of the area (direct ties) who collaborate with the firm and their contacts in the area (indirect ties). The results of expanding the variables of interest are shown in Table A4. Overall, connections linked by non-stars are more strongly related to entry than connections linked by stars. However, the social capital of star scientists within the area does appear of higher influence to firms' decision on entry. Although star scientists may be less available in time and mind to reflect on research opportunities with the firm, the firm may be able to abuse her reputation among the scientists in their own research to obtain information or establish new connections. However, based on this supplementary analysis, it is difficult to draw strong conclusion. Hence, I encourage further research on the topic possibly determining stardom within the focal area and also examining stardom in second line.

3.8.4 Examining inter-firm variance

Within the main regressions, I choose to include dummy variables to control for firm specific elements related to their openness for the exploration of new scientific areas, like firms'

[31] The star identification process is explained in the second chapter of this dissertation.

corporate management, culture or strategy. The regression presented in Table A5 does not include these dummies and shows robustness of the findings for the insider connections. Though, the variable for the outsider-connected share of area's scientists loses its significance in this regression setup. This finding suggest that firms' who are more strongly connected to areas by outsiders are likely to share some firm-level attribute. One possibility of such an attribute is its geographical location (Owen-Smith and Powell 2004b). Firms positioned within innovation hubs may have a more explorative culture while being member of a dense social network between local scientists.

Additionally, the regression results specify that firms with a higher scientific relatedness in their innovations, higher number and success of completed projects, and higher scientific diversity are more likely to enter a new research area. Belderbos (2016) emphasized the importance of firms' scientific absorptive capacity, approximated by their focus on basic research, to enable good communication and knowledge transfer with academic collaboration partners. Hence, in line with their study, Table A6 presents a split sample analysis. Firms are considered absorptive for scientific research when their publications of the past four years contain more basic research (CHI-journal classification level 3 and 4) than the average within the sample. It appears that social connections influence high and low absorptive firms in a similar way.

3.8.5 Long-term effect of social connections

It can be argued from a tertius uingens perspective that indirect ties form a step to the generation of (additional) direct connections: the direct tie redirects the firm to one of her contacts (Quintane & Carnabuci, 2016). If this is the case, indirect ties may influence entry in advance of direct ties. Hence, in this robustness check I explain entry in year t by the firms' social connections established through collaboration in t-6 to t-3. Table A7 presents the results which are consistent with the main model. Hence, I find no support for the tertius uingens perspective.

3.9 DISCUSSION

The results provide support for the three hypotheses, though also present some additional insights. In general, enlarging firm's connectivity to a research area, by increasing the share of the area's scientists which are members of the firm's social network, increases firms' likelihood on entering the research area. However, when taking a more detailed view, it appears that this higher connectivity is only positively related to firms' likelihood to enter the area when connected to the firm by an insider of the area. Higher connectivity to area scientists to which the firm is only connected by an outsider of the area appears unrelated to firms entry decision.

Due to frictions and the lack of broker ability, I expected that an outsider would not communicate all information present at her contacts in the area and would not recognize and process the information as well as an insider. However, the complete lack, and in some models even negative relationship between these outsider connections and firms' entry into the research area is still surprising. In the following, I speculate on five possible explanations for the surprisingly low coefficient, which opens directions for future research.

First, the outsider may not only communicate incomplete information but also incorrect information. When information does not correspond with other information gathered by the firm, uncertainty may demotivate firms to enter the area. Second, firms may not be aware of the lack and incompleteness of the information they receive from outsiders as they are themselves not knowledgeable within the area. Hence, firms which are connected to the area's scientific community by outsiders may experience a false feeling of being informed while they are actually missing out on large entry opportunities within the area. As the formation of absorptive capacity is a costly process, past research has indeed indicated that firms often ineffectively limit their search to the local area and mainly rely on their friends, in-group members and similar minded individuals for advice (Song *et al.* 2018; Mcdonald and Westphal 2003). Third, in line with the study of low performing CEO's who lack to respond with strategic decisions, firms may not want to explore new research areas and are searching for confirmation from their network that sustained focus on their current research areas is optimal. In such a situation, collaborators who are not active in the focal area may provide this confirmation and so increase firms' likelihood to not enter the area. Furthermore, outsiders may opportunistically demotivate firms to take on opportunities in other areas in the hope of safeguarding their financial support by the firm. While university-industry collaborations often represent an important source of funding for academics (D'Este and Perkmann, 2011; Perkman*n et al*, 2013), outsiders may want to avoid that their industrial partners invest in another research area than its own. Finally, the establishment of outsider connections may only form a first step in gathering information on a research opportunity. The outsider may provide a first, though important, signal that there are some advancements within the area. Although she cannot provide details, she may follow the tertius iungens strategy (Quintane and Carnabuci 2016) and facilitate the formation of direct connections to the area. Within the regression analysis, this would imply that outsider connections to the area's community might only influence entry in the long run after they have resulted in additional information gathering through more reliable connections.

The results further indicate that firm's likelihood to enter the research area increases when there is a right balance between direct and indirect ties among the area scientists who are connected to the firm by an insider. Even after controlling for the average quality of the directly tied insider connections, the results provide support for the idea of Burt (1992) that indirect ties are an effective strategy for keeping abreast of new opportunities. However, I need to make three important nuances. First, we only find this effect for indirect ties that are mediated by an insider in the research area. This condition implies that firms, who hope to expand their activity across scientific areas, may benefit from collaboration with multidisciplinary scientists. Second, the results are positioned within the particular context of entry into new scientific areas. As communication, information disclosure, knowledge transfer and idea sharing are key attributes of the scientific community, other contexts characterized by secrecy and contractual constraints may benefit more from direct ties as they may need close interaction to benefit from complementary knowledge and trust. Finally, it could be argued that, although information may suffice to recognize an opportunity and find the confidence to act upon it, firms are likely to require deeper understanding from embedded knowledge to perform within an area. The supplementary analysis on publication performance after entry supports this reasoning.

3.10 CONCLUSION

Within high-tech industries, scientific research capabilities are considered key for firms' competitive advantage (Deeds, Decarolis, and Coombs 2000; Cassiman, Veugelers, and Zuniga 2008; Gambardella 1992). However, to sustain this advantage firms need to stay informed on new opportunities so that they can adjust the direction of their scientific activity accordingly *(Deeds, Decarolis, and Coombs 2000; Siciliano, Welch, and Feeney 2018; Iain M Cockburn et al.* 2000). Considering the strong relationship between collaboration with academic scientists and firms' innovation performance (Belderbos, Carree, and Lokshin 2004; Furman and MacGarvie 2009), I questioned in this study whether firms' connections to the scientific community of a research area may be related to firms' decision to start own research activity in the area. I posit that social connections may keep the firm abreast on opportunities within that area or may provide important information after interest in the area has been established by the firm.

The results signal, that even when including firm dummies and controls for the area's overall attractiveness, the social connections that the firm has to the area's scientific community does influence their likelihood to establish their first publication in the area. Moreover, the particular

formation of these connections appears of crucial importance. While I expected firms to be more likely to enter a research area when establishing higher connectivity to its scientists, the results indicate that only those connections established through collaboration with an insider in the area have a positive influence. Connections that are established through outsiders of the research area appear insignificant for firms' probability to enter. Additionally, the findings provide a first indication that information of a scientist in the area can be obtained through indirect ties to that scientist without substantial loss of its informative value for the firm under two conditions. First, the mediator of the information is also an active contributor of the area, i.e. an insider, enabling her to communicate complementary knowledge alongside the received information to the firm. Second, the information exchange takes place within the scientific community where knowledge sharing, disclosure, open communication and critique form components of daily life.

The study emphasizes the importance of affinity to a research area for absorbing and communication upon it. While past research on indirect ties has mainly ignored the role of the broker, the results indicate that outsiders, lacking affinity to the area, are ineffective and even detrimental access points to its information. Some studies have acknowledge the importance of the broker or intermediate (Belderbos *et al.* 2016a; Fleming *et al.* 2007; Quintane & Carnabuci 2016; Salman & Saives 2005) and have identified benefits of proximity between the broker and receiver of information (Belderbos, Gilsing, and Suzuki 2016b; Guan and Liu 2016). Complementary, this study signals that affinity with the knowledge space of the information may also form an important attribute of a good broker.

I also contribute to the debate on the relative importance of direct and indirect ties. Studies analyzing new contexts are important for enhancing our insight in information exchange within social networks (Ahuja 2000). While the context of the scientific community is characterized by highly complex and tacit knowledge, it appears that information, in its broad sense, can be effectively exchanged across direct ties under a capable broker and that information may already be highly valuable for guiding the entry behavior of firms. The study, hence, dampens the high enthusiasm on direct and close interaction within the university-industry literature (Zucker and Darby, 1996; Zucker *et al*., 1998; Fabrizio, 2009) by showing that indirectly accessing scientists may also be of value for firms.

As our results emphasizes the importance of the communicator of information or alternatively the first-line contacts, the study supports the ideas of Burt (2007) that direct ties form the most

important components within the social network. However, as we do find that indirect insider connections have an additional stimulating power on entry in comparison to direct insider connection, I do not agree with Burt's (Burt 2007) related statement that the investigation of indirect ties only provides limited additional explanatory value and may be ignored within the analysis of social networks. In contrast, I believe that this study and others (Shah, Levin and Cross, 2018) which investigate the conditions under which indirect ties can be influential for the performance and decisions of actors, are of high importance. Thanks to the low relational costs of indirect ties, they may form, under the right condition, an effective strategy for firms to process the exponentially growing amount of information available (M. Granovetter 1985b; Burt 1997; Burt 2004; Mckelvey and Rake 2016).

Moreover, by taking into account the quality of firms' collaboration partners within the area, this study signals that beside human capital, which forms the core subject of investigation in the university-industry literature (Zucker & Darby 1996; Fuller & Rothaermel 2012; Hess & Rothaermel 2011), also social capital is a valuable attribute of academic collaboration partners. The results namely signal that, when considering the exploration of new research areas, the contacts of a firms' collaboration partner can form important sources of information for a firm and that this information can, under the conditions mentioned, be effectively transferred by the partner.

Notwithstanding these contributions, the study has its limitations. First, there is a possible concern of endogeneity within the data. It is, namely, unclear whether firms engage in certain collaborative partnerships after having the tendency to enter in a particular research area. However, as the scientific community is highly open for collaborative initiatives and industrial partners often provide additional funding or the opportunity to see the materialization of discoveries, it could be expected that firms with a clear tendency would be able to immediately start-up a collaboration in the area itself. Related, it can be argued that outsider connection have no influence on firms entry behaviour as they are not established with the intent to gather information on the focal area. However, the fact that firms do not have the intention to utilize these connections as sources of information already signals that firms question their ability to provide information or doubt the credibility of the information received. Second, while I refer to effectiveness of strategies this only implies an increased probability on reaction considering the number of connections established. I cannot make statements on the financial effectiveness of the analyzed strategies as I am not aware of data on budgets assigned to explorative activities, especially not in regards to one particular area. Third, I apply a specific taxonomy of disease

areas that may have its influence on the results. Moreover, I only examined tie formation while the literature has also discussed the importance of overall network structure (C. Phelps, Heidl, and Wadhwa 2012). Finally, the social connections have been measured through co-publications while scientists and firms may of course establish social connections in more informal ways like conference attendance or informal visits. Fabrizio (2009) for instance mentioned that Novartis builds connections to university scientists by sponsoring weekly seminars on site, where university scientists discuss their work. Additionally, Tzabbar (2009) has discussed how recruitment may influence firms technological positioning.

The inclusion of network structure and scientist mobility would form interesting avenues for further research. Moreover, as the level of analysis is restricted to a particular research area, it would be interesting to analyze characteristics of an area and its knowledge on the importance of social network connections. Past findings have, for instance, indicated the importance of direct ties for sharing complex and tacit knowledge (Darby and Zucker 2003; Sorenson, Rivkin, and Fleming 2006). Additionally, I would like to encourage further research on brokerage as the findings signal high importance of the intermediate. Research could examine brokers' knowledge diversity, overlap with the firm or network position within the area. I also suggest future researchers to define a moment of firm's established interest into the research area possibly by first scientific reference or NPR to a publication of the area. By observing the social network from this moment, network formation and network change can be examined under the intention of entry. Finally, it would be interesting to analyze the network dynamics after entry into the area. Such work may provide evidence of the tertius uingens strategy, and show possible remaining benefits of indirect ties for research performance. Finally,

3.11 APPENDIX

Table 3.A1. Cloglog regression predicting firm's probability to enter a research area persistently (at least 5 publication in coming 8 years)

	Cloglog Persistent entry	
	Coef.	(z) p
VARIABLES OF INTEREST		
Outsider connected share of area's scientists	-27.92	(-4.53) 0.00
Insider connected share of area's scientists	18.37	(4.40) 0.00
No. of insider connected indirect ties/ no. of insider connected direct ties	0.03	(6.71) 0.00
FIRM CONTROLS		
Log(R&D $_{t-1, \text{million \$}}$)	0.38	(3.02) 0.00
Scientific relatedness of innovations	0.02	(1.99) 0.05
Number of completed commercialization projects	0.00	(0.70) 0.48
Share of successfully completed projects	-0.17	(-0.21) 0.84
Scientific diversity	-0.04	(-2.63) 0.01
No. of publications within overarching research area/R&D	0.41	(1.85) 0.06
Basic research focus of publications in overarching research area	0.29	(2.40) 0.02
AREA CONTROLS		
Number of active contributors in area	0.01	(6.52) 0.00
Competition in research area	7.66	(7.89) 0.00
Dummy(growing research area)	0.72	(4.48) 0.00
Basic research focus of research area	0.00	(3.22) 0.00
Breadth of research area	0.01	(2.30) 0.02
Age of research area	0.30	(2.99) 0.00

Coefficients in bold represent relations with a p-value < 0.05. Standard errors are clustered at the firm-area level.

Table 3.A1 continued.

	Cloglog persistent entry	
	Coef.	(z) p
FIRM-AREA CONTROLS		
Distance of research area to firm	0.00	(0.03) 0.98
Average quality of insider connected direct ties	0.10	(2.31) 0.02
DUMMIES		
Year dummies	Yes	
Firm dummies	No	
Overarching research area dummies	Yes	
Constant	-13.66	(-10.41) 0.00
N	171524	
LL	-3341.23	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A2. Descriptive statistics on firms' publications in the research area in the first two years since entry

VARIABLES	**MEAN**	**SD**	**MIN**	**MAX**
Total number of publications in area in first two years	1.527	1.097	1.000	18.000
… Share in collaboration with academia	0.867	0.313	0.000	1.000
... Share in collaboration with previous				
Outsider connections	.006	.054	0	1
Direct insider connections	.565	.458	0	1
Indirect insider connections	.021	.097	0	1
Total number of collaboration partners in first two years	9.809	8.861	1.000	135.000
…Share to which firm had				
An outsider connection	0.001	0.011	0.000	0.222
A direct insider connection	0.264	0.272	0.000	1.000
An indirect insider connection	0.058	0.232	0.000	1.000
Dummy whether firm published all research of the first two years alone	0.101	0.301	0.000	1.000

Table 3.A3. two-stage Poisson regression predicting firm's publication count in the first two years of entering the research area

	All publications		**Single publications**		**U-I collaborative publications**	
	Coef.	(z) p	Coef.	(z) p	Coef.	(z) p
VARIABLES OF INTEREST						
Outsider connected share of area's scientists	-4.39	(-3.46) 0.00	1.50	(0.34) 0.73	-3.52	(-2.38) 0.02
Insider connected share of area's scientists	3.46	(3.79) 0.00	-3.18	(-1.05) 0.29	4.63	(4.54) 0.00
No. of insider connected indirect ties/ no. of insider connected direct ties	0.00	(0.52) 0.60	0.00	(0.21) 0.83	0.00	(0.49) 0.63
FIRM CONTROLS						
Log(R&D $_{\text{t-1, million \$}}$)	0.03	(0.69) 0.49	0.09	(0.58) 0.56	-0.08	(-1.76) 0.08
Scientific diversity	-0.00	(-0.84) 0.40	-0.00	(-0.18) 0.86	0.01	(1.54) 0.12
No. of publications within overarching research area/R&D	0.09	(1.32) 0.19	0.11	(0.40) 0.69	0.04	(0.55) 0.58
Basic research focus of publications in overarching research area	0.10	(3.06) 0.00	0.28	(2.59) 0.01	0.07	(1.75) 0.08
RESEARCH AREA CONTROLS						
Number of active contributors in area	0.00	(1.85) 0.06	0.00	(0.66) 0.51	0.00	(1.82) 0.07
Basic research focus of research area	0.05	(2.21) 0.03	0.26	(3.36) 0.00	0.02	(0.56) 0.57
Breadth of research area	0.00	(0.11) 0.91	-0.00	(-1.42) 0.16	0.00	(0.39) 0.69
FIRM-AREA CONTROLS						
Distance of research area to firm	0.02	(1.05) 0.29	0.04	(0.77) 0.44	0.01	(0.49) 0.63
Average quality of insider connected direct ties	0.02*	(1.68) 0.09	-0.01	(-0.26) 0.79	0.03	(3.17) 0.00
DUMMIES						
Year dummies	Yes		Yes		Yes	
Firm dummies	Yes		Yes		Yes	
Overarching research area dummies	Yes		Yes		Yes	
Mills ratio	-0.10	(-5.49) 0.00	-0.14	(-1.99) 0.05	-0.07	(-3.44) 0.00
Constant	0.66	(2.20) 0.03	-0.96	(-0.82) 0.41	0.48	(1.43) 0.15
LL	-3876.97		-1561.56		-3459.15	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A4. Cloglog panel regression with random effects predicting firm's probability to enter a research area controlling for first-stage strength of the connections

	Model 1		**Model 2**	
	Coef.	(z) p	Coef.	(z) p
VARIABLES OF INTEREST				
Outsider connected share of area's scientists	1.23	(0.54) 0.59	2.55	(1.17) 0.24
Insider connected share of area's scientists	12.99	(7.27) 0.00	11.51	(7.79) 0.00
No. of insider connected indirect ties/ no. of insider connected direct ties	0.03	(9.61) 0.00	0.02	(8.59) 0.00
Average strength of first stage ties to outsider connected scientists of area	-0.79	(-12.37) 0.00	-0.72	(-11.22) 0.00
Average strength of first stage ties to insider connected scientists of area	0.88	(9.56) 0.00		
Average strength of first stage ties to direct insider connected scientists of area			0.11	(4.51) 0.00
Average strength of first stage ties to indirect insider connected scientists of area			0.63	(9.32) 0.00
FIRM CONTROLS				
Log(R&D $_{\text{t-1, million \$}}$)	0.25	(4.36) 0.00	0.24	(4.05) 0.00
Scientific relatedness of innovations	0.01	(1.83) 0.07	0.01	(1.90) 0.06
Number of completed commercialization projects	-0.00	(-1.41) 0.16	-0.00	(-1.71) 0.09
Share of successfully completed projects	0.13	(0.40) 0.69	0.14	(0.41) 0.68
Scientific diversity	0.00	(0.11) 0.91	0.00	(0.39) 0.70
No. of publications within overarching research area/R&D	0.44	(4.17) 0.00	0.38	(3.39) 0.00
Basic research focus of publications in overarching research area	0.13	(2.20) 0.03	0.13	(2.10) 0.04
RESEARCH AREA CONTROLS				
Number of active contributors in area	0.01	(11.91) 0.00	0.01	(11.31) 0.00
Competition in research area	3.93	(23.18) 0.00	3.97	(22.18) 0.00
Dummy(growing research area)	0.28	(4.89) 0.00	0.28	(4.87) 0.00

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A4 continued.

	Model 1		Model 2	
	Coef.	(z) p	Coef.	(z) p
Basic research focus of research area	0.11	(2.80) 0.01	0.11	(2.82) 0.00
Breadth of research area	0.00	(2.00) 0.05	0.00	(1.86) 0.06
Age of research area	0.01	(3.53) 0.00	0.01	(3.54) 0.00
FIRM-AREA CONTROLS				
Distance of research area to firm	-0.05	(-1.66) 0.10	-0.05	(-1.51) 0.13
Average quality of insider connected direct ties	0.12	(6.50) 0.00	0.13	(6.68) 0.00
DUMMIES				
Year dummies	Yes		Yes	
Firm dummies	Yes		Yes	
Overarching research area dummies	Yes		Yes	
Constant	-8.49	(-17.53) 0.00	-8.60	(-17.67) 0.00
N	207207		207207	
LL	-12113.65		-12005.41	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A5. Cloglog panel regression with random effects predicting firm's probability to enter a research area controlling for stardom in first line

	Coef.	(z) p
VARIABLES OF INTEREST		
Outsider connected share of area's scientists linked by non-star	7.71	(2.88) 0.00
Outsider connected share of area's scientists linked by star	-15.46	(-3.88) 0.00
Insider connected share of area's scientists linked by non-star	20.48	(8.93) 0.00
Insider connected share of area's scientists linked by star	14.38	(3.43) 0.00
No. indirect ties of connected insider non-star/ no. of insider non-stars connected directly	-0.00	(-0.46) 0.65
No. indirect ties of connected insider star/ no. of insider stars connected directly	0.02	(13.10) 0.00
FIRM CONTROLS		
Log(R&D $_{\text{t-1, million \$}}$)	0.23	(4.05) 0.00
Scientific relatedness of innovations	0.01	(2.29) 0.02
Number of completed commercialization projects	-0.00	(-1.03) 0.30
Share of successfully completed projects	0.19	(0.56) 0.58
Scientific diversity	-0.01	(-0.78) 0.43
No. of publications within overarching research area/R&D	0.46	(4.69) 0.00
Basic research focus of publications in overarching research area	0.12	(1.96) 0.05
RESEARCH AREA CONTROLS		
Number of active contributors in area	0.01	(11.26) 0.00
Competition in research area	3.94	(23.25) 0.00
Dummy(growing research area)	0.29	(4.96) 0.00
Basic research focus of research area	0.14	(3.63) 0.00
Breadth of research area	0.00	(2.32) 0.02
Age of research area	0.01	(3.33) 0.00

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A5 continued.

	Coef.	(z) p
FIRM-AREA CONTROLS		
Distance of research area to firm	-0.04	(-1.30) 0.19
Average quality of insider connected direct ties	No	
DUMMIES		
Year dummies	Yes	
Firm dummies	No	
Overarching research area dummies	Yes	
Constant	-8.61	(-17.67) 0.00
N	207207	
LL	-12346.90	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A6. Cloglog random effects regression predicting firms' probability to enter a research area without firm dummies

	Coef.	(z) p
VARIABLES OF INTEREST		
Outsider connected share of area's scientists	0.71	(0.33) 0.74
Insider connected share of area's scientists	16.00	(6.77) 0.00
No. of insider connections indirectly tied/ no. of insider connections directly tied	0.03	(11.97) 0.00
FIRM CONTROLS		
Log(R&D $_{t-1, \text{million \$}}$)	0.32	(12.56) 0.00
Scientific relatedness of innovations	0.01	(3.38) 0.00
Number of completed commercialization projects	0.00	(2.88) 0.00
Share of successfully completed projects	0.42	(2.78) 0.01
Scientific diversity	0.05	(10.82) 0.00
No. of publications within overarching research area/R&D	0.61	(7.30) 0.00
Basic research focus of publications in overarching research area	0.22	(3.84) 0.00
RESEARCH AREA CONTROLS		
Number of active contributors in area	0.01	(12.60) 0.00
Competition in research area	3.90	(22.74) 0.00
Dummy(growing research area)	0.29	(5.06) 0.00
Basic research focus of research area	0.12	(2.95) 0.00
Breadth of research area	0.00	(1.91) 0.06
Age of research area	0.01	(3.34) 0.00

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A6 continued.

	Coef.	(z) p
FIRM-AREA CONTROLS		
Distance of research area to firm	-0.05	(-1.45) 0.15
Average quality of insider connected direct ties	0.14	(6.82) 0.00
DUMMIES		
Year dummies	Yes	Yes
Firm dummies	No	No
Overarching research area dummies	Yes	Yes
Constant	-10.94	(-38.62) 0.00
N	207207.00	
LL	-12711.82	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A7. Cloglog panel regression with random effects predicting firm's probability to enter a research area, split sample according to the scientific absorptive capacity of the firm

	Above average absorptive firms		**Below average Absorptive firms**	
	Coef.	(z) p	Coef.	(z) p
VARIABLES OF INTEREST				
Outsider connected share of area's scientists	-5.29	(-1.32) 0.19	-4.37	(-1.58) 0.11
Insider connected share of area's scientists	15.64	(8.12) 0.00	13.96	(4.77) 0.00
No. indirect insider connections/ no. of direct insider connections	0.03	(5.84) 0.00	0.03	(10.13) 0.00
FIRM CONTROLS				
Log(R&D $_{t-1, \text{million \$}}$)	-0.03	(-0.25) 0.80	0.36	(4.78) 0.00
Scientific relatedness of innovations	0.01	(0.71) 0.48	0.01	(1.39) 0.16
Number of completed commercialization projects	-0.02	(-2.78) 0.01	-0.00	(-0.26) 0.80
Share of successfully completed projects	0.04	(0.06) 0.95	0.38	(0.91) 0.36
Scientific diversity	-0.02	(-1.00) 0.32	0.01	(0.59) 0.55
No. of publications within overarching research area/R&D	0.36	(3.29) 0.00	2.87	(5.18) 0.00
Basic research focus of publications in overarching research area	0.06	(0.59) 0.56	0.12	(1.62) 0.11
RESEARCH AREA CONTROLS				
Number of active contributors in area	0.01	(8.24) 0.00	0.01	(10.29) 0.00
Competition in research area	4.22	(12.28) 0.00	3.72	(19.49) 0.00
Dummy(growing research area)	0.34	(3.25) 0.00	0.28	(4.16) 0.00
Basic research focus of research area	-0.00	(-1.73) 0.08	0.00	(3.44) 0.00
Breadth of research area	0.00	(0.63) 0.53	0.01	(3.34) 0.00
Age of research area	0.17	(2.28) 0.02	0.11	(2.53) 0.01

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A7 continued.

	Above average absorptive firms		**Below average Absorptive firms**	
	Coef.	(z) p	Coef.	(z) p
FIRM-AREA CONTROLS				
Distance of research area to firm	-0.10	(-1.55) 0.12	-0.04	(-0.91) 0.36
Average quality of insider connected direct ties	0.16	(7.51) 0.00	0.11	(4.20) 0.00
DUMMIES				
Year dummies	Yes		Yes	
Firm dummies	Yes		Yes	
Overarching research area dummies	Yes		Yes	
Constant	-0.14	(-0.49) 0.63	-0.20	(-0.81) 0.42
N	57615		148579	
LL	-3601.46		-8637.82	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A8. Cloglog random effects regression predicting firms' probability to enter a research area explained by social connections established from t-6 to t-3

	Model 4	
	Coef.	(z) p
VARIABLES OF INTEREST		
Outsider connected share of area's scientists	-1.38	(-0.59) 0.55
Insider connected share of area's scientists	8.07	(8.74) 0.00
No. of insider connections indirectly tied/ no. of insider connections directly tied	0.02	(7.06) 0.00
FIRM CONTROLS		
Log(R&D $_{\text{t-1, million \$}}$)	0.13*	(1.68) 0.09
Scientific relatedness of innovations	0.02	(2.90) 0.00
Number of completed commercialization projects	-0.01	(-3.03) 0.00
Share of successfully completed projects	0.27	(0.57) 0.57
Scientific diversity	-0.00	(-0.34) 0.73
No. of publications within overarching research area/R&D	0.43	(4.18) 0.00
Basic research focus of publications in overarching research area	0.19	(2.87) 0.00
RESEARCH AREA CONTROLS		
Number of active contributors in area	0.01	(10.01) 0.00
Competition in research area	3.67	(18.49) 0.00
Dummy(growing research area)	0.31	(4.72) 0.00
Basic research focus of research area	0.18	(4.31) 0.00
Breadth of research area	0.00	(2.48) 0.01
Age of research area	0.01	(4.08) 0.00

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

Table 3.A8 continued.

	Coef.	(z) p
FIRM-AREA CONTROLS		
Distance of research area to firm	-0.22	(-0.92) 0.36
Average quality of insider connected direct ties	0.05	(3.03) 0.00
DUMMIES		
Year dummies	Yes	Yes
Firm dummies	Yes	Yes
Overarching research area dummies	Yes	Yes
Constant	-7.33	(-11.02) 0.00
N	138326	
LL	-8512.77	

Coefficients in bold represent relations with a p-value < 0.05.
Standard errors are clustered at the firm-area level.

General Conclusion

This dissertation has provided additional insights on the role of science within the pharmaceutical industry by investigating different performance indicators along the development process and by taking a more detailed look at the research performed and the partners selected. In this conclusion, I will first reflect on the contribution and managerial implications of my findings in respect to the current productivity crisis within pharmaceuticals. Second, I will acknowledge the limitations of the datasets used and of the studies' research scope. Finally, I will make suggestions for future research.

SCIENCE AND THE PRODUCTIVITY CRISIS

As explained within the introduction of this dissertation, the pharmaceutical industry has been suffering under decreasing returns of R&D investments since the nineties. Discussions on this so-called productivity crisis have resulted in three major concerns: the increasing challenge of capturing opportunities (Danzon, Nicholson and Pereira, 2005; Pammolli, Magazzini and Riccaboni, 2011; Moors, Cohen and Schellekens, 2014; Crispeels, Willems and Scheerlinck, 2018), the lack of fundamental and translational understanding (Kola and Landis, 2004; Cockburn, 2007) and the growing organizational barriers between research and development (Cockburn, 2007; Arora et al., 2009). In general, the findings of this dissertation support prior research on the beneficial role of science and, more precisely, of in-house basic research and collaboration with academia for the innovation performance of (bio-)pharmaceutical companies. Moreover, all chapters indicate that the effectiveness of these scientific efforts can be further increased when appropriately tailored to the specific concerns mentioned within the productivity debate.

Chapter 1 signals that invented molecules with a scientific background are considered more valuable candidates for drug development than those lacking scientific insight. This finding suggest that firms believe that 'good' drug candidates build upon scientific research and that failure in development is related to the lack thereof. However, combined with the opinion in the literature that too much candidates in development lack scientific understanding, this result may have three implications: (1) firms should select even more strongly on the candidates' scientific background, (2) there are just insufficient candidates with a proper scientific

background, as the low-hanging fruit has been picked, or (3) the disaggregation of the industry has disabled the industry's ability to assign a drug candidate to their most suited developer. In respect to the latter, the chapter's findings suggest that the firm that embeds the fundamental understanding of the drug candidate is more likely to take-on development. While irrational considerations for extended investment of drug candidate may form an explanation, the analyses show greater support for the fact that the embedder of fundamental knowledge may be the most-suited developer. This finding corresponds with the concern that the division of labor across research and development may be detrimental for productivity as organizational barriers interrupt the transfer of basic understanding from discovery to development.

The second chapter focused on collaboration with academic star scientists. When pertaining to the increasing difficulty of health problems, these collaborations may be regarded of special importance for firms to establish the needed scientific capabilities. The study's findings indicate that, conditional on no involvement of an internal star scientist, the effectiveness of collaboration with academic stars can be improved by involving the star scientist in applied research with the firm. In absence of this involvement, an internal star scientist may take-over this translational responsibility within the joint-basic research. These findings further emphasize the innovation benefit of combining basic and applied scientific understanding in research projects and, hence, vertical integration of research and development. Moreover, like Chapter 1, this study shows the difficulty of crossing organizational borders within research projects. Additional to the applied understanding within project teams, the results stress the importance that an innovation premium can only be realized when the star scientists are also dedicated to their industrial partner. Beside the avoidance of knowledge spillovers to rivals, I posit that this dedication is key for establishing trustworthy and frequent interaction that facilitates knowledge transfer.

In the last chapter of this dissertation, I investigated whether collaboration with academia may help firms in their quest for new research opportunities. The results show that firms are more likely to enter new research areas when they can establish insider connections to the area. Outsider connections, however, appear irrelevant to firms' probability on entry. If, as suggested in the discussion of the productivity crisis, the available research opportunities are more complex challenges requiring large investments in new capabilities, outsiders may be unlikely to see the complete picture and may not be seen as credible information sources for such a risky decision. Firms are likely to need complementary knowledge of an insider to correctly frame the pieces of information available and identify those that are still missing.

IMPLICATIONS

Overall, this dissertation provides three important implications for managers who aim to increase the effectiveness of their research and development expenditures. First, our results further support the importance of scientific understanding for companies within the pharmaceutical industry. While prior is has emphasized its role in the generation of new inventions (Arora, Belenzon, & Patacconi, 2015; Gambardella, 1992; Kelchtermans, Leten, & Belderbos, 2013), Chapter 1 of this dissertation also indicates its role in investment decisions for the further development of these inventions.

Second, while scientific understanding can be embedded within the firm or can be obtained through collaboration or licensing, it is important for firms to recognize that it often consists out of tacit and complex knowledge that is difficult to transfer across organizational and institutional borders. Consequently, collaboration with academic star scientists may be negatively related to firms' innovation performance. The basic-research background of a drug candidate appears irrelevant for the take-up decision of firms other than the embedder of the basic research. Finally, outsiders to a scientific research area seem not to influence the entry of firms into the area.

The findings finally conclude that, although difficult, firms can effectively cross organizational and institutional borders under the right contingencies. First, boundary crossing knowledge transfer may be facilitated by codification of the knowledge or by having assistance in interpreting and framing the knowledge. The first chapter showed that when internal research results in - more codified and application-specific - applied publications, it can influence the development expectations of firms in the market for technology. Also, the other chapters indicate that knowledge can be more easily obtained from external sources like collaboration partners, or indirect connections when the firm is assisted in the interpretation and translation. Star academic scientists are more suited collaboration partners to firms when also performing applied research with the partner firm and information on a new research area strongly benefits from communication by scientists active in the area. Second, this dissertation indicated that boundaries can be more easily crossed when close relationships between two sides is established. Collaboration between firms and academic star scientists aimed at acquiring fundamental understanding may benefit from dedication of the collaboration partner. Drug candidates are more likely to cross the boundary between research and development when both are performed within the same company. Information on an unknown research area is more

strongly considered by firms when the area's boundary is crossed by the first-line connection of the firm (the insider).

LIMITATIONS

While this dissertation provides insights to the productivity discussion, it also inevitably has its limitations in examining the role of science within the pharmaceutical sector. First, the studies rely heavily on publication data. Although, the publication activity of pharmaceutical companies is high, it can be argued that not all research projects and interactions with academia are covered by the data analyzed. Moreover, the publication behavior of firms may have changed over the years (Arora et al. 2009). However, as every study includes year dummies, a change in the industry-level appreciation of publishing should not influence the results. Second, the studies only reflect on the period of 1995 to 2002 due to data limitations which limits the number of observations, especially within Chapter 1, and restricts the possibilities of analyzing long-term effects or changing trends. The latter, especially, limits the contribution of this dissertation to the productivity debate as no comparison is possible with pre-'crisis' periods. Third, while the data utilized by this dissertation is divers and highly detailed, it lacks financial information like research budgets, development costs, or staff and material expenditures. While all studies control for the research and development expenditures covered, effectiveness statements are merely based on marginal effects per engaged collaboration or established connection. Fourth, the dissertation's databases report on the innovation and research output of the most R&D intensive firms of the pharmaceutical industry. Hence, its conclusions only reflect on the industry's larger players and can not be used to formulate implications for small players like start-ups or niche players. They may still benefit from a strong informal and internalized link to universities limiting the need for dedication or translation of collaborating star scientists. Also, they may not have the means or ambitions to develop and market their drug candidates internally which would considerably change the take-up decision investigated in the firs chapter. Finally, this study ignores team effects. Among scientists, I did not take into account lab membership and, among firms, I did not consider consortiums or incubators. However, I believe these team effects would not influence my main findings as I often indirectly observe them by consistent mentioning of lab heads on all publications of the lab and by controlling for alliances or taking aggregated values for drug candidates owned by multiple firms.

This dissertation only examines the role of science within the pharmaceutical industry, however also other industries are highly involved with science or could benefit from higher involvement. While I acknowledge the strong focus on pharmaceuticals within my studies, I do believe that the findings can be translated to other industries in which product development is a costly and long process. In such an environment, development through experimentation is inferior to the development of fundamental understanding leading to more directed applied research. Also in these industries, I believe that firms' innovation performance would benefit from collaboration with dedicated and translational star scientists, product candidates with stronger scientific background and social connections to new research areas.

FUTURE RESEARCH DIRECTIONS

While the science-industry link already received a lot of attention in past research and this dissertation has provided additional insights, I still encourage further research on the topic. First, future research may want to focus on the characteristics and social position of gatekeepers within companies. As Chapter 3 indicated the large importance of good brokerage for establishing information exchange and stimulating reactions, it may be interesting to analyze how external information and knowledge is distributed within companies. Especially, the increased size and multidisciplinary character of research projects can make it difficult, if not impossible, for one person to, not only, absorb all knowledge, but also, transfer it to the R&D team. Further, it may be interesting to investigate the characteristics of the research area like basicness and cognitive distance, in more detail. Second, Chapter 2 signaled that the configuration and tasks of internal and external scientists within teams might strongly impact the innovation benefit of research projects. I would like to encourage future research on this team configuration across institutions, possibly through the examination of author positions. Third, the scientists and firms that I have interviewed often referred to differences between clinical and university scientists. It may be interesting to investigate whether these differences also imply other collaboration benefits and possibly different contingencies for effective collaboration. Fourth, when analyzing the development of drug candidates in Chapter 1, I focused on the perspective of the patent owner. However, the perspective of the lead company may also form an interesting subject for future research. Not only, could these studies investigate whether take-up into development is greater when the lead company is acquainted with the scientific background of the drug candidate, though may also link this pre-licensing

knowledge to knowledge-gathering efforts after licensing like collaborative development (Crispeels, Willems, and Scheerlinck 2018).

References

Adams, Christopher P, and Van V Brantner. 2006. 'Estimating the Cost of New Drug Development: Is It Really 802 Million Dollars?' *Health Affairs* 25 (2). Project HOPE - The People-to-People Health Foundation, Inc.: 420–28. doi:10.1377/hlthaff.25.2.420.

Adams, James D. 1990. 'Fundamental Stocks of Knowledge and Productivity Growth'. *Journal of Political Economy* 98 (4): 673–702.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C Stein. 2008. 'Academic Freedom, Private-Sector Focus, and the Process of Innovation'. *Journal of Economics*. Vol. 39.

Agrawal, Ajay. 2006. 'Engaging the Inventor: Exploring Licensing Strategies for University Inventions and the Role of Latent Knowledge'. *Strategic Management Journal* 27 (1): 63–79. doi:10.1002/smj.508.

Ahuja, Gautam. 2000a. 'Collaboration Networks, Structural Holes, and Innovation : A Longitudinal Study'. *Administrative Science Quarterly* 45 (3): 425–55. doi:10.2307/2667105.

———. 2000b. 'Collaboration Networks , Structural Holes , and Innovation : A Longitudinal Study'. *Administrative Science Quarterly* 45 (3): 425–55. doi:10.2307/2667105.

Ahuja, Gautam, C Lampert, and E Novelli. 2013. 'The Second Face of Appropriability: Generative Appropriability and Its Determinants'. *Academy of Management Review* 38 (2): 248–69.

Almeida, P., J. Hohberger, and P. Parada. 2011. 'Individual Scientific Collaborations and Firm-Level Innovation'. *Industrial and Corporate Change* 20 (6): 1571–99. doi:10.1093/icc/dtr030.

Almeida, Paul, Jan Hohberger, and Pedro Parada. 2011. 'Individual Scientific Collaborations and Firm-Level Innovation'. *Industrial and Corporate Change* 20 (6): 1571–99. doi:10.1093/icc/dtr030.

Antons, David, and Frank T Piller. 2015. 'Opening the Black Box of "Not Invented Here": Attitudes, Decision Biases, and Behavioral Consequences' 29 (2): 193–217.

Aral, Sinan, and Marshall Van Alstyne. 2011. 'The Diversity-Bandwidth Trade-Off'. *Source: American Journal of Sociology* 117 (1). The University of Chicago Press AJS: 90–171. doi:10.1086/661238.

Arora, Ashish, Sharon Belenzon, and Andrea Patacconi. 2015. 'Killing the Golden Goose? The Decline of Science in Corporate R&D'.

———. 2017. 'The Decline of Science in Corporate R&D'. *Strategic Management Journal*, no. June 2017: 3–32. doi:10.1002/smj.2693.

Arora, Ashish, Sharon Belenzon, and Lia Sheer. 2017. 'Back to Basics: Why Firms Invest in Research?'

Arora, Ashish, and Alfonso Gambardella. 1990. 'Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology'. *The Journal of Industrial Economics* 38 (4). Wiley: 361. doi:10.2307/2098345.

———. 1994. 'Evaluating Technolological Information and Utilizing It* Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology'. *Journal of Economic Behavior & Organization* 24: 91–114.

———. 2010. 'Ideas for Rent: An Overview of Markets for Technology'. *Industrial and Corporate Change* 19 (3): 775–803. doi:10.1093/icc/dtq022.

Arora, Ashish, Alfonso Gambardella, Laura Magazzini, and Fabio Pammolli. 2009. 'A Breath of Fresh Air? Firm Type, Scale, Scope, and Selection Effects in Drug Development'. *MANAGEMENT SCIENCE* 55 (10): 1638–53. doi:10.1287/mnsc.1090.1055.

Arrow, K. 1962. *Economic Welfare and the Allocation of Resources for Invention*. *The Rare and Direction of Inventive Activity*. Princeton University Press.

Arts, Sam, Francesco Paolo Appio, and Bart Van Looy. 2013. 'Inventions Shaping Technological Trajectories:

Do Existing Patent Indicators Provide a Comprehensive Picture?' *Scientometrics* 97: 397–419. doi:10.1007/s11192-013-1045-1.

Audretsch, David B., and Paula E. Stephan. 1996. 'Company-Scientist Locational Links: The Case of Biotechnology'. *American Economic Review* 86 (3): 641–52. doi:10.2307/2118217.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. 'Superstar Extinction'. *Quaterly Journal of Economics* 125 (May): 549–89.

Azoulay, Pierre, Toby E. Stuart, and Yanbo Wang. 2014. 'Matthew: Effect or Fable?' *Management Science* 60 (August): 92–109. doi:doi:10.1287/mnsc.2013.1755.

Baba, Yasunori, Naohiro Shichijo, and Silvia Rita Sedita. 2009. 'How Do Collaborations with Universities Affect Firms' Innovative Performance? The Role of "Pasteur Scientists" in the Advanced Materials Field'. *Research Policy* 38 (5): 756–64. doi:10.1016/j.respol.2009.01.006.

Bakker, Jurrien, Dennis Verhoeven, Lin Zhang, and Bart Van Looy. 2016. 'Patent Citation Indicators : One Size Fits All ?' *Scientometrics* 106 (1): 187–211.

Balachandran, Sarath, and Exequiel Hernandez. 2018. 'Networks and Innovation: Accounting for Structural and Institutional Sources of Recombination in Brokerage Triads'. *Organization Science* 29 (1): 80–99. doi:10.1287/orsc.2017.1165.

Banal-Estanol, Albert, Ines Macho-Stadler, and David Perez-Castrillo. 2014. 'Endogenous Matching in University-Industry Collaboration: Theory and Empirical Evidence from the UK'.

Banal-estañol, Albert, Inés Macho-stadler, and David Pérez-castrillo. 2017. 'Endogenous Matching in University-Industry Collaboration : Theory and Empirical Evidence from the United Kingdom Endogenous Matching in University-Industry Collaboration : Theory and Empirical Evidence from the United Kingdom'. *Management Science, Articles in Advance*, no. March: 1–18.

Barrenho, Eliana, Marisa Miraldo, and Peter C Smith. 2013. 'The Determinants of Attrition in Drug Development: A Duration Analysis'.

Belderbos, René, Martin Carree, and Boris Lokshin. 2004. 'Cooperative R&D and Firm Performance'. *Research Policy* 33 (10). North-Holland: 1477–92. doi:10.1016/J.RESPOL.2004.07.003.

Belderbos, René, Bruno Cassiman, Dries Faems, Bart Leten, Bart Van Looy, and Ku Leuven. 2014. 'Co-Ownership of Intellectual Property: Exploring the Value-Appropriation and Value-Creation Implications of Co-Patenting with Different Partners'. *Research Policy* 43: 841–52. doi:10.1016/j.respol.2013.08.013.

Belderbos, René, Victor A Gilsing, and Shinya Suzuki. 2016a. 'Direct and Mediated Ties to Universities: " Scientific " Absorptive Capacity and Innovation Performance of Pharmaceutical Firms'. *Strategic Organization* 14 (1): 32–52. doi:10.1177/1476127015604734.

———. 2016b. 'Direct and Mediated Ties to Universities: "Scientific" Absorptive Capacity and Innovation Performance of Pharmaceutical Firms'. *Strategic Organization* 14 (1): 32–52. doi:10.1177/1476127015604734.

Blind, Knut, Katrin Cremers, and Elisabeth Mueller. 2009. 'The Influence of Strategic Patenting on Companies' Patent Portfolios'. *Research Policy* 38 (2). North-Holland: 428–36. doi:10.1016/J.RESPOL.2008.12.003.

Blundell, R, R Griffith, and J Van Reenen. 1999. 'Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms'. *The Economic Journal* 66 (3): 529–54. doi:10.1111/1467-937X.00097.

Blundell, R, R Griffith, and J Vanreenen. 1995. 'Dynamic Count Data Models of Technological Innovation'. *The Economic Journal* 105 (429): 333–44.

Boyack, Kevin W., Michael Patek, Lyle H. Ungar, Patrick Yoon, and Richard Klavans. 2014. 'Classification of Individual Articles from All of Science by Research Level'. *Journal of Informetrics* 8 (1). Elsevier Ltd: 1–12. doi:10.1016/j.joi.2013.10.005.

Bruneel, Johan, Pablo D'Este, and Ammon Salter. 2010. 'Investigating the Factors That Diminish the Barriers to University–industry Collaboration'. *Research Policy* 39 (7). Elsevier B.V.: 858–68. doi:10.1016/j.respol.2010.03.006.

Burt, Ronald S. 2004. ‘Structural Holes and Good Ideas’. *American Journal of Sociology* 110 (2): 349–99.

Burt, Ronald S., and Jennifer Merluzzi. 2016. ‘Network Oscillation’. *Academy of Management Discoveries* 2 (4): 368–91. doi:10.5465/amd.2015.0108.

Burt, Ronald S. 1992. ‘The Social Structure of Competition’. In *Networks and Organizations*, 57–91. Cambridge: harvard university press.

———. 1997. ‘The Contingent Value of Social Capital.’ *Administrative Science Quarterly* 42 (2): 339–65. doi:10.2307/2393923.

———. 2000. ‘The Network Structure of Social Capital’. *Research in Organizational Behaviour* 22: 345–423.

———. 2007. ‘Secondhand Brokerage: Evidence on the Importance of Local Structure for Managers’. *The Academy of Management Journal* 50 (1): 119–48.

Callaert, Julie, Maikel Pellens, and Bart Van Looy. 2013. ‘Sources of Inspiration? Making Sense of Scientific References in Patents’. *Scientometrics* 98 (3): 1617–29. doi:10.1007/s11192-013-1073-x.

Cameron, C.A., and P.K. Trivedi. 2009. ‘Microeconometrics Using Stata’. Stata Press.

Campbell, J.J. 2005. *Understanding Pharma: A Primer on How Pharmaceutical Companies Really Work.* Raleigh: Pharmaceutical Institute.

Cassiman, Bruno, and Reinhilde Veugelers. 2006. ‘In Search of Complementarity in Innovation Strategy: Internal R1D and External Knowledge Acquiqisition’. *Management Science* 52 (1): 68–82. doi:10.1287/mnsc.l050.0470.

Cassiman, Bruno, Reinhilde Veugelers, and Pluvia Zuniga. 2008. ‘In Search of Performance Effects of (in)Direct Industry Science Links’. *Industrial and Corporate Change* 17 (4): 611–46. doi:10.1093/icc/dtn023.

Cattani, G., S. Ferriani, M. M. Mariani, and S. Mengoli. 2013. ‘Tackling the “Galacticos” Effect: Team Familiarity and the Performance of Star-Studded Projects’. *Industrial and Corporate Change* 22 (6): 1629–62. doi:10.1093/icc/dtt001.

Ceccagnoli, Marco, and Lin Jiang. 2013. ‘The Cost of Integrating External Technologies: Supply and Demand Drivers of Value Creation in the Markets for Technology’. *Strategic Managment Journal* 34: 404–25. doi:10.1002/smj.

Chesbrough, Henry, and Chen Eric L. 2013. ‘Recovering Abandoned Compounds Through Expanded External IP Licensing.: EBSCOhost’. *California Management Review* 55 (4): 83–103.

Chiou, Jing-yuan, Laura Magazzini, Fabio Pammolli, and Massimo Riccaboni. 2012. ‘The Value of Failure in Pharmaceutical R & D’.

Chiou, Jing-Yuan, Laura Magazzini, Fabio Pammolli, and Massimo Riccaboni. 2016. ‘Learning from Successes and Failures in Pharmaceutical R&D’. *Journal of Evolutionary Economics* 26: 271–90. doi:10.1007/s00191-015-0439-z.

Cockburn, Iain M. 2006. ‘Is the Pharmaceutical Industry in a Productivity Crisis?’ *Innovation Policy and the Economy* 7 (January). MIT Press: 1–32. doi:10.1086/ipe.7.25056188.

Cockburn, Iain M. 2004. ‘The Changing Structure Of The Pharmaceutical Industry’. *Health Affairs* 23 (1): 10–22. doi:10.1377/hlthaff.23.1.10.

Cockburn, Iain M, and Rebecca M Henderson. 1998. ‘Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery’. *The Journal of Industrial Economics* 46 (2): 157–82.

Cockburn, Iain M, Rebecca M Henderson, Scott Stern, and Ali Sina Önder. 2000. ‘Untangling the Origins of Competitive Advantage’. *Strategic Management Journal* 21: 1123–45.

Cockburn, Iian M. 2004. ‘The Changing Structure Of The Pharmaceutical Industry’. *Health Affairs* 23 (1): 10–22. doi:10.1377/hlthaff.23.1.10.

Cockburn, Im. 2007. *Is the Pharmaceutical Industry in a Productivity Crisis? Innovation Policy and the Economy, Volume 7*. Vol. 7.

Cockburn, Iain M, and Rebecca M Henderson. 2001. 'Publicly Funded Science and the Productivity of the Pharmaceutical Industry'. In *Innovation Policy and the Economy, Volume 1*, 1–34.

Cohen, Wesley M, and Daniel A Levintbal. 1990. 'Absorptive Capacity : A New Perspective on Learning and Innovation'.

Cohen, Wesley M, and Daniel A Levinthal. 1989. 'Innovation and Learning: The Two Faces of R&D' 99 (397): 569–96.

Coleman, James S. 1988a. 'Social Capital in the Creation of Human Capital'. *American Journal of Sociology* 94 (1988): S95. doi:10.1086/228943.

Coleman, James S. 1988b. 'Social Capital in the Creation of Human Capital Social Capital in the Creation of Human Capital' AJS Volume 94 Supplement S95-S120 S95'. *Source: American Journal of Sociology* 94: 95–120.

Colen, Relinde Maria, Rene Belderbos, Stijn Kelchtermans, and Bart Leten. 2018. 'Reaching for the Stars : The Contingent Performance Effects of Basic Research Collaboration'.

Colin, Terry, and Karen Taylor. 2018. 'A New Future for R&D | Deloitte UK'. *Deloitte*. https://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/measuring-return-from-pharmaceutical-innovation.html.

Colyvas, Jeannette, Michael Crow, Annetine Gelijns, Roberto Mazzoleni, Richard R, Nathan Rosenberg, Bhaven N Sampat, Institutefor International Studies, and Richard R Nelson. 2002. 'How Do University Inventions Get into Practice ?' *Management Science, Special Issue on University Entrepreneurship and Technology Transfer* 48 (1): 61–72.

Cook, David, Dearg Brown, Robert Alexander, Ruth March, Paul Morgan, Gemma Satterthwaite, and Menelas N Pangalos. 2014. 'Lessons Learned from the Fate of AstraZeneca's Drug Pipeline: A Five-Dimensional Framework'. *Nature Publishing Group* 13. doi:10.1038/nrd4309.

Crispeels, Thomas, Jurgen Willems, and Ilse Scheerlinck. 2018. 'Public-Private Collaboration in Drug Development: Boosting Innovation or Alleviating Risk?' *Public Management Review* 20 (2): 273–92. doi:10.1080/14719037.2017.1302247.

D'Este, Pablo, and Markus Perkmann. 2011. 'Why Do Academics Engage with Industry? The Entrepreneurial University and Individual Motivations'. *Journal of Technology Transfer* 36: 316–39. doi:10.1007/s10961-010-9153-z.

Danzon, Patricia M, Sean Nicholson, and Nuno Sousa Pereira. 2005. 'Productivity in Pharmaceutical–biotechnology R&D: The Role of Experience and Alliances'. *Journal of Health Economics* 24 (24): 317–39. doi:10.1016/j.jhealeco.2004.09.006.

Darby, Michael R, and Lynne G Zucker. 2003. 'Grilichesian Breakthroughs: Inventions of Methods of Inventing and Firm Entry in Nanotechnology'. 9825. NBER Working Paper. Cambridge.

Deeds, David L, DONA Decarolis, and Joseph Coombs. 2000. 'Dynamic Capabilities and New Product Development in High Technology Ventures: An Empirical Analysis of New Biotechnology Firms'. *Journal of Business Venturing* 15 (3). Elsevier: 211–29. doi:10.1016/S0883-9026(98)00013-5.

DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. 2016. 'Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs'. *Journal of Health Economics* 47. Elsevier B.V.: 20–33. doi:10.1016/j.jhealeco.2016.01.012.

Dimasi, Joseph A, Ronald W Hansen, and Henry G Grabowski. 2003. 'The Price of Innovation: New Estimates of Drug Development Costs'. *Journal of Health Economics* 22 (2): 151–85. doi:10.1016/S0167-6296(02)00126-1.

DiMasi, Joseph A, Ronald W Hansen, and Henry G Grabowski. 2003. 'The Price of Innovation: New Estimates of Drug Development Costs'. *Journal of Health Economics* 22 (2): 151–85. doi:10.1016/S0167-6296(02)00126-1.

Dobson, Paul D., and Douglas B. Kell. 2008. 'Carrier-Mediated Cellular Uptake of Pharmaceutical Drugs: An Exception or the Rule?' *Nature Reviews Drug Discovery* 7 (3). Nature Publishing Group: 205–20. doi:10.1038/nrd2438.

Eichler, Hans-Georg, Eric Abadie, June M. Raine, and Tomas Salmonson. 2009. 'Safe Drugs and the Cost of Good Intentions'. *New England Journal of Medicine* 360 (14). Massachusetts Medical Society : 1378–80. doi:10.1056/NEJMp0900092.

Elfenbein, Daniel W., and Josh Lerner. 2012. 'Exclusivity, Contingent Control Rights, and the Design of Internet Portal Alliances'. *Journal of Law, Economics, and Organization* 28 (1): 45–76. doi:10.1093/jleo/ewp035.

Eliëns, Ramon, Katrin Eling, Sarah Gelper, and Fred Langerak. 2018. 'Rational Versus Intuitive Gatekeeping: Escalation of Commitment in the Front End of NPD'. *Journal of Product Innovation Management* 0 (0): 1–18. doi:10.1111/jpim.12452.

Fabrizio, Kira R. 2009. 'Absorptive Capacity and the Search for Innovation'. *Research Policy* 38 (2). Elsevier B.V.: 255–67. doi:10.1016/j.respol.2008.10.023.

Faems, Dries, Maddy Janssens, and Bart Van Looy. 2010. 'Managing the Co-Operation-Competition Dilemma in R&D Alliances: A Multiple Case Study in the Advanced Materials Industry'. *Creativity and Innovation Management* 19 (1): 3–22. doi:10.1111/j.1467-8691.2010.00546.x.

Fleming, Lee, Santiago Mingo, and David Chen. 2007. 'Collaborative Brokerage, Generative Creativity, and Creative Success'. *Administrative Science Quarterly* 52: 443–75.

Fleming, Lee, and Olav Sorenson. 2004. 'Science as a Map in Technological Search'. *Strategic Management Journal* 25 (89): 909–28. doi:10.1002/smj.384.

Foray, Dominique, and Francesco Lissoni. 2010. 'University Research and Public-Private Interaction'. In *Economics of Innovation*, 276–308.

Fuller, Anne W., and Frank T. Rothaermel. 2012. 'When Stars Shine: The Effects of Faculty Founders on New Technology Ventures'. *Strategic Entrepreneurship Journal* 6 (3): 220–35. doi:10.1002/sej.1140.

Furman, J. L., and M. MacGarvie. 2009. 'Academic Collaboration and Organizational Innovation: The Development of Research Capabilities in the US Pharmaceutical Industry, 1927-1946'. *Industrial and Corporate Change* 18 (5). Oxford University Press: 929–61. doi:10.1093/icc/dtp035.

Gambardella, Alfonso. 1992. 'Competitive Advantages from In-House Scientific Research: The US Pharmaceutical Industry in the 1980s'. *Research Policy* 21 (5): 391–407. doi:10.1016/0048-7333(92)90001-K.

Gambardella, Alfonso, P. Giuri, and A. Luzzi. 2007. 'The Market for Patents in Europe'. *Research Policy* 36 (8): 1163–83.

Gambardella, Alfonso, Dietmar Harhoff, and Bart Verspagen. 2008. 'The Value of European Patents'. *European Management Review* 5 (2). Blackwell Publishing Ltd: 69–84. doi:10.1057/emr.2008.10.

Ghosh, Anindya, and Lori Rosenkopf. 2015. 'PERSPECTIVE—Shrouded in Structure: Challenges and Opportunities for a Friction-Based View of Network Research'. *Organization Science* 26 (2): 622–31. doi:10.1287/orsc.2014.0922.

Gilsing, Victor, Bart Nooteboom, Wim Vanhaverbeke, Geert Duysters, and Ad van den Oord. 2008. 'Network Embeddedness and the Exploration of Novel Technologies: Technological Distance, Betweenness Centrality and Density'. *Research Policy* 37 (10): 1717–31. doi:10.1016/j.respol.2008.08.010.

Gilsing, Victor, Bart Nooteboom, Wim Vanhaverbeke, Geert Duysters, and Ad Van Den Oord. 2008. 'Network Embeddedness and the Exploration of Novel Technologies: Technological Distance, Betweenness Centrality and Density'. *Research Policy* 37: 1717–31. doi:10.1016/j.respol.2008.08.010.

Gittelman, Michelle. 2016. 'The Revolution Re-Visited: Clinical and Genetics Research Paradigms and the Productivity Paradox in Drug Discovery'. *Research Policy* 45 (8). North-Holland: 1570–85. doi:10.1016/J.RESPOL.2016.01.007.

Gittelman, Michelle, and Bruce Kogut. 2003. 'Does Good Science Lead to Valuable Knowledge ? Biotechnology Firms and the Evolutionary Logic of Citation Patterns', no. 1973.

Gould, Roger V, and Roberto M Fernandez. 1989. 'Structures of Mediation: A Formal Approach to Brokerage in Transaction Networks'. *Sociological Methodology*. Vol. 19.

Granovetter, Mark. 1985a. ‘Economic Action and Social Structure: The Problem of Embeddedness’. *American Journal of Sociology* 91 (3): 481–510. doi:10.1086/228311.

———. 1985b. ‘Economic Action and Social Structure: The Problem of Embeddedness’. *American Journal of Sociology* 91 (3): 481–510. doi:10.1086/228311.

———. 2013. ‘Economic Action and Social Structure : The Problem of Embeddedness ’’ 91 (3): 481–510.

Granovetter, Mark S. 1973. ‘The Strength of Weak Ties’. *American Journal of Sociology* 78 (6): 1360–80.

Grigoriou, K., and F. T. Rothaermel. 2013. *Structural Microfoundations of Innovation: The Role of Relational Stars*. *Journal of Management*. Vol. 40. doi:10.1177/0149206313513612.

Groysberg, B., J. T. Polzer, and H. a. Elfenbein. 2011. ‘Too Many Cooks Spoil the Broth: How High-Status Individuals Decrease Group Effectiveness’. *Organization Science* 22 (3): 722–37. doi:10.1287/orsc.1100.0547.

Groysberg, Boris, Linda-Eling Lee, and Ashish Nanda. 2008. ‘Can They Take It With Them? The Portability of Star Knowledge Workers’ Performance’. *Management Science* 54 (7): 1213–30. doi:10.1287/mnsc.1070.0809.

Groysberg, Boris, Ashish Nanda, and Nitin Nohria. 2004. ‘The Risky Business of Hiring Stars.’ *Harvard Business Review* 82 (5): 92–100, 151.

Guan, Jiancheng, and Na Liu. 2016. ‘Exploitative and Exploratory Innovations in Knowledge Network and Collaboration Network: A Patent Analysis in the Technological Field of Nano-Energy’. *Research Policy* 45: 97–112. doi:10.1016/j.respol.2015.08.002.

Hagedoorn, John, and Ning Wang. 2012. ‘Is There Complementarity or Substitutability between Internal and External R&D Strategies?’ *Research Policy* 41 (6). Elsevier B.V.: 1072–83. doi:10.1016/j.respol.2012.02.012.

Hall, B., A. Jaffe, and M. Trajtenberg. 2005. ‘Market Value and Patent Citations’. *Rand Journal of Economics* 36 (1): 16–38.

Hamilton, Kimberly S. 2003. ‘Subfield and Level Classification of Journals’.

Hausman, Jerry, and Daniel Mcfadden. 1984. ‘Specification Tests for the Multinomial Logit Model’. Vol. 52.

Hay, Michael, David W Thomas, John L Craighead, Celia Economides, and Jesse Rosenthal. 2014. ‘Clinical Development Success Rates for Investigational Drugs’ 32 (1).

Hess, Andrew M., and Frank T. Rothaermel. 2011. ‘When Are Assets Complementary? Star Scientists, Strategic Alliances, and Innovation in the Pharmaceutical Industry’. *Strategic Management Journal* 32 (8). John Wiley & Sons, Ltd.: 895–909. doi:10.1002/smj.916.

Hess, Andrew, and Frank T. Rothaermel. 2012. ‘Intellectual Human Capital and the Emergence of Biotechnology: Trends and Patterns, 1974–2006’. *IEEE Transactions on Engineering Management* 59 (1): 65–76. doi:10.1109/TEM.2010.2082550.

Hicks, Diana M., Phoebe A. Isard, and Ben R. Martin. 1996. ‘A Morphology of Japanese and European Corporate Research Networks’. *Research Policy* 25 (3): 359–78. doi:10.1016/0048-7333(95)00830-6.

Jaffe, Adam B. 1989. ‘Real Effects of Aca (Jemic Research’. *The American Economic Review* 79 (5): 957–71.

Jayaraj, Sebastian, and Michelle Gittelman. 2018. ‘Scientific Maps and Innovation: Impact of the Human Genome on Drug Discovery’. In *DRUID Society Conference 2018*, 1–56.

Jenkins, Stephen P. 2005. ‘Survival Analysis’. In *Essex Summer School Course ‘Surival Analysis’*, 1–133.

Jensen, Richard A., Jerry G. Thursby, and Marie C. Thursby. 2003. ‘Disclosure and Licensing of University Inventions: “The Best We Can Do with the S**t We Get to Work With”’. *International Journal of Industrial Organization* 21: 1271–1300. doi:10.1016/S0167-7187(03)00083-3.

Kaplan, S., Fiona Murray, and Rebecca Henderson. 2003. ‘Discontinuities and Senior Management: Assessing the Role of Recognition in Pharmaceutical Firm Response to Biotechnology’. *Industrial and Corporate Change* 12 (2). Oxford University Press: 203–33. doi:10.1093/icc/12.2.203.

Kehoe, Rebecca R., and Daniel Tzabbar. 2015. 'Lighting the Way or Stealing the Shine? An Examination of the Duality in Star Scientists' Effects on Firm Innovative Performance'. *Strategic Management Journal* 36: 709–27. doi:10.1002/smj.

Kelchtermans, S., and R. Veugelers. 2011. 'The Great Divide in Scientific Productivity: Why the Average Scientist Does Not Exist'. *Industrial and Corporate Change* 20 (1): 295–336. doi:10.1093/icc/dtq074.

Kelchtermans, Stijn, Bart Lete, and Rene Belderbos. 2013. 'Do Firms Benefit from Investing in Basic Research? An Empirical Investigation for Pharmaceutical Firms'. In *Innovation and Creativity*, edited by De Beule, 155–72.

Kelchtermans, Stijn, and Reinhilde Veugelers. 2013. 'Top Research Productivity and Its Persistence: Gender as a Double-Edged Sword' 95 (March): 273–85.

Kim, Hui Kwon, Seonwoo Min, Myungjae Song, Soobin Jung, Jae Woo Choi, Younggwang Kim, Sangeun Lee, Sungroh Yoon, and Hyongbum (Henry) Kim. 2018. 'Deep Learning Improves Prediction of CRISPR–Cpf1 Guide RNA Activity'. *Nature Biotechnology*, January. Nature Publishing Group. doi:10.1038/nbt.4061.

Klueter, Thomas, L Felipe Monteiro, and Denise R Dunlap. 2017. 'Standard vs. Partnership-Embedded Licensing_ Attention and the Relationship between Licensing and Product Innovations'. doi:10.1016/j.respol.2017.07.013.

Kola, Ismail, and John Landis. 2004. 'Can the Pharmaceutical Industry Reduce Attrition Rates?' *Nature Reviews: Drug Discovery* 3 (August): 1–5.

Lahiri, Nandini. 2010. 'Geographic Distribution of R&D Activity: How Does It Affect Innovation Quality?' *Academy of Management Journal* 53 (5): 1194–1209. doi:10.5465/AMJ.2010.54533233.

Laudel, Grit. 2002. 'What Do We Measure by Co-Authorships?' *Research Evaluation* 11 (1): 3–15. doi:10.3152/147154402781776961.

Laursen, K., M.I. Leone, and S. Torrisi. 2010. 'Technological Exploration through Licensing: New Insights from the Licensee's Point of View'. *Industrial and Corporate Change* 19 (3): 871–97.

Leone, M.I., and Toke Reichstein. 2012. 'Licensing-in Fosters Rapid Invention! The Effect of the Grant-Back Clause and Technological Unfamiliarity'. *Strategic Managment Journal* 33 (8): 965–85.

Leten, Bart, Rene Belderbos, and Bart Van Looy. 2016. 'Entry and Technological Performance in New Technology Domains: Technological Opportunities, Technology Competition and Technological Relatedness'. *Journal of Management Studies* 53 (8): 1257–91. doi:10.1111/joms.12215.

Lewison, Grant, and Guillermo Paraje. 2004. 'The Classification of Biomedical Journals by Research Level'. *Scientometrics* 60 (2): 145–57. doi:10.1023/B:SCIE.0000027677.79173.b8.

Lichtenberg, Frank R. 2005. 'The Impact of New Drug Launches on Longevity: Evidence from Longitudinal, Disease-Level Data from 52 Countries, 1982–2001'. *International Journal of Health Care Finance and Economics* 5 (1): 47–73.

Lichtenthaler, Ulrich. 2011. 'The Evolution of Technology Licensing Management: Identifying Five Strategic Approaches'. *R&D Management* 41 (2). Wiley/Blackwell (10.1111): 173–89. doi:10.1111/j.1467-9310.2011.00635.x.

Lim, Kwanghui. 2004. 'The Relationship between Research and Innovation in the Semiconductor and Pharmaceutical Industries (1981-1997)'. *Research Policy* 33 (2): 287–321. doi:10.1016/j.respol.2003.08.001.

Lissoni, Francesco, and Fabio Montobbio. 2015. 'Guest Authors or Ghost Inventors? Inventorship Attribution in Academic Science'. *Evaluation Review* 39 (1): 19–45.

Lissoni, Francesco, Fabio Montobbio, and Lorenzo Zirulia. 2013. 'Inventorship and Authorship as Attribution Rights: An Enquiry into the Economics of Scientific Credit'. *Journal of Economic Behavior and Organization* 95. Elsevier B.V.: 49–69. doi:10.1016/j.jebo.2013.08.016.

Lotka, Alfred J. 1926. 'The Frequency Distribution of Scientific Productivity'. *Journal of the Washington Academy of Sciences* 16 (12). Washington Academy of Sciences: 317–24. doi:10.2307/24529203.

Luo, Xiaowei Rose, Kenneth W. Koput, and Walter W. Powell. 2009. 'Intellectual Capital or Signal? The Effects of Scientists on Alliance Formation in Knowledge-Intensive Industries'. *Research Policy* 38 (8): 1313–25. doi:10.1016/j.respol.2009.06.001.

Malik, Tariq. 2011. 'Vertical Alliance and Vertical Integration for the Inflow of Technology and New Product Development in the Pharmaceutical Industry'. *Technology Analysis & Strategic Management* 23 (8). Routledge : 851–64. doi:10.1080/09537325.2011.604154.

Malva, Antonio Della, Stijn Kelchtermans, Bart Leten, and Reinhilde Veugelers. 2015. 'Basic Science as a Prescription for Breakthrough Inventions in the Pharmaceutical Industry'. *Journal of Technology Transfer* 40 (4). Springer US: 670–95. doi:10.1007/s10961-014-9362-y.

Malva, Antonio Della, Bart Leten, Stijn Kelchtermans, and Reinhilde Veugelers. 2015. 'Basic Science as a Prescription for Technological Breakthroughs in the Pharmaceutical Industry'. *Journal of Technology Transfer* 40: 670–95. doi:10.1007/s10961-014-9362-y.

Mansfield, Edwin. 1995. 'Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing'. *The Review of Economics and Statistics* 77 (1): 55–62.

———. 1998. 'Academic Research and Industrial Innovation: An Update of Empirical Findings'. *Research Policy* 26 (7–8). North-Holland: 773–76. doi:10.1016/S0048-7333(97)00043-7.

March, James G., Herbert A. (Herbert Alexander) Simon, and Harold Steere. Guetzkow. 1993. *Organizations*. Blackwell.

Martinez, C. 2011. 'Patent Families: When Do Different Definitions Really Matter?' *Scientometrics* 86 (1): 39–63.

Maxwell, Robert A., and Shohreh B. Eckhardt. 1990. *Drug Discovery : A Casebook and Analysis*. Humana Press.

Mcdonald, Michael L, and James D Westphal. 2003. 'Getting by with the Advice of Their Friends: CEOs' Advice Networks and Firms' Strategic Responses to Poor Performance'. *Source: Administrative Science Quarterly*. Vol. 48.

Mckelvey, Maureen, and Bastian Rake. 2016. 'Product Innovation Success Based on Cancer Research in the Pharmaceutical Industry: Co-Publication Networks and the Effects of Partners'. *Industry and InnovatIon* 23 (5): 383–406. doi:10.1080/13662716.2016.1150157.

Melin, G., and O. Persson. 1996. 'Studying Research Collaboration Using Co-Authorships'. *Scientometrics* 36 (3): 363–77. doi:10.1007/BF02129600.

Meyer, M., and O. Persson. 1998. 'Nanotechnology-Interdisciplinarity, Patterns of Collaboration and Differences in Application'. *Scientometrics* 42 (2). Kluwer Academic Publishers: 195–205. doi:10.1007/BF02458355.

Meyer, Martin. 2006. 'Are Patenting Scientists the Better Scholars?' *Research Policy* 35 (10): 1646–62. doi:10.1016/j.respol.2006.09.013.

Mindruta, Denisa. 2013. 'value creation in university-firm research collaborations : a matching approach'. *Strategic Management Journal* 34: 644–65. doi:10.1002/smj.

Moors, Ellen H M, Adam F Cohen, and Huub Schellekens. 2014. 'Towards a Sustainable System of Drug Development'. *Drug Discovery Today* 19 (11): 1711–20. doi:10.1016/j.drudis.2014.03.004.

Morgan, Paul, Piet H. Van Der Graaf, John Arrowsmith, Doug E. Feltner, Kira S. Drummond, Craig D. Wegner, and Steve D.A. Street. 2012. 'Can the Flow of Medicines Be Improved? Fundamental Pharmacokinetic and Pharmacological Principles toward Improving Phase II Survival'. *Drug Discovery Today* 17 (9–10). Elsevier Current Trends: 419–24. doi:10.1016/J.DRUDIS.2011.12.020.

Murray, Fiona. 2004. 'The Role of Academic Inventors in Entrepreneurial Firms: Sharing the Laboratory Life'. *Research Policy* 33 (4): 643–59. doi:10.1016/j.respol.2004.01.013.

Murray, Fiona, and Scott Stern. 2007. 'Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis'. *Journal of Economic Behavior and Organization* 63 (4): 648–87. doi:10.1016/j.jebo.2006.05.017.

Narin, Francis, Kimberly S Hamilton, and Dominic Olivastro. 1997. 'The Increasing Linkage between U.S. Technology and Public Science'. *ELSEVIER Research Policy* 26: 317–30.

Natalicchio, a., a. Messeni Petruzzelli, and a.C. Garavelli. 2014. 'A Literature Review on Markets for Ideas: Emerging Characteristics and Unanswered Questions'. *Technovation* 34 (2). Elsevier: 65–76. doi:10.1016/j.technovation.2013.11.005.

Nelson, Richard R. 1959. 'The Simple Economics of Basic Scientific Research'. *The Journal of Political Economy* 67 (3): 297–306.

Oncurious. 2015. 'Oncurious NV Appoints Prof DR Peter Carmelies as Its Chief Scientific Strategy Advisor'. *Press Release*.

Owen-Smith, Jason, and Walter W. Powell. 2004a. 'Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community'. *Organization Science* 15 (1): 5–21.

———. 2004b. 'Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community'. *Organization Science* 15 (1). INFORMS: 5–21. doi:10.1287/orsc.1030.0054.

Owen-Smith, Jason, and Walter W Powell. 2004c. 'Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community'. *Organization Science* 15 (1): 5–21. doi:10.1287/orsc.1030.0054.

Pammolli, Fabio, Laura Magazzini, and Massimo Riccaboni. 2011. 'The Productivity Crisis in Pharmaceutical R&D'. *Nature Reviews Drug Discovery* 10 (6). Nature Publishing Group: 428–38. doi:10.1038/nrd3405.

Paul, Steven M., Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg, and Aaron L. Schacht. 2010. 'How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge'. *Nature Reviews Drug Discovery* 9 (3). Nature Publishing Group: 203–14. doi:10.1038/nrd3078.

Peck, Richard W, Dennis W Lendrem, Iain Grant, B Clare Lendrem, and John D Isaacs. 2015. 'Why Is It Hard to Terminate Failing Projects in Pharmaceutical R&D?' *Nature Publishing Group* 14: 663–64. doi:10.1038/nrd4725.

Perkmann, Markus, Valentina Tartari, Maureen McKelvey, Erkko Autio, Anders Broström, Pablo D'Este, Riccardo Fini, et al. 2013a. 'Academic Engagement and Commercialisation: A Review of the Literature on University–industry Relations'. *Research Policy* 42 (2). Elsevier B.V.: 423–42. doi:10.1016/j.respol.2012.09.007.

———. 2013b. 'Academic Engagement and Commercialisation: A Review of the Literature on University–industry Relations'. *Research Policy* 42 (2). Elsevier B.V.: 423–42. doi:10.1016/j.respol.2012.09.007.

———. 2013c. 'Academic Engagement and Commercialisation: A Review of the Literature on University–industry Relations'. *Research Policy* 42 (2). Elsevier B.V.: 423–42. doi:10.1016/j.respol.2012.09.007.

Perry-Smith, Jill E., and Pier Vittorio Mannucci. 2016. 'From Creativity to Innovation: The Social Network Drivers of the Four Phases of the Idea Journal'. *Academy of Management Review* 42 (1): 53–79. doi:10.5465/amr.2014.0462.

Phelps, C., R. Heidl, and a. Wadhwa. 2012. *Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda*. *Journal of Management*. Vol. 38. doi:10.1177/0149206311432640.

Phelps, Corey, Ralph Heidl, and Anu Wadhwa. 2012. 'Knowledge, Networks, and Knowledge Networks'. *Journal of Management* 38 (4). SAGE PublicationsSage CA: Los Angeles, CA: 1115–66. doi:10.1177/0149206311432640.

Phelps, Corey, Ralph Heidl, Anu Wadhwa, and Hec Paris. 2012. 'Knowledge, Networks, and Knowledge Networks : A Review and Research Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda'. *Journal of Management* 38 (4): 1115–66. doi:10.1177/0149206311432640.

Pisano, Gary P. 1997a. 'R&D Performance, Collaborative Arrangements and the Market for Know-How: A Test of the "Lemons" Hypothesis in Biotechnology'. *SSRN Electronic Journal*. doi:10.2139/ssrn.41980.

———. 1997b. ‘R&D Performance, Collaborative Arrangements and the Market for Know-How: A Test of the “Lemons” Hypothesis in Biotechnology’. *Available at SSRN: Http://Ssrn.Com/Abstract=41980*. doi:10.2139/ssrn.41980.

Plewa, Carolin, Nisha Korff, Thomas Baaken, and Greg Macpherson. 2013. ‘University-Industry Linkage Evolution: An Empirical Investigation of Relational Success Factors’. *R&D Management* 43 (4): 365–80. doi:10.1111/radm.12021.

Polidoro, Frangisgo. 2013. ‘The competitive implications of certifications: the effects of scientific and regulatory certifications on entries into new technical fields’. *® Academy of Management Journal* 56 (2): 597–827. doi:10.5465/amj.2011.0459.

Powell, Walter W., Kenneth W. Koput, and Laurel Smith-Doerr. 1996. ‘Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology’. *Administrative Science Quarterly* 41 (1): 116–45.

Quintane, Eric, and Gianluca Carnabuci. 2016. ‘How Do Brokers Broker? Tertius Gaudens, Tertius Iungens, and the Temporality of Structural Holes’. *Organization Science* 27 (6): 1343–60. doi:10.1287/orsc.2016.1091.

Radauer, A., and T. Dudenbostel. 2013. ‘PATLICE Survey: Survey on Patent Licensing Activities by Patenting Firms’. In *European Commission, Directorate & General for Research and Innovation*, 26114.

Rafols, Ismael, Michael M. Hopkins, Jarno Hoekman, Josh Siepel, Alice O’Hare, Antonio Perianes-Rodríguez, and Paul Nightingale. 2014. ‘Big Pharma, Little Science?. A Bibliometric Perspective on Big Pharma’s R&D Decline.’ *Technological Forecasting and Social Change* 81. Elsevier Inc.: 22–38. doi:10.1016/j.techfore.2012.06.007.

Rappert, Brian, Andrew Webster, and David Charles. 1999. ‘Making Sense of Diversity and Reluctance: Academic–industrial Relations and Intellectual Property’. *Research Policy* 28: 873–90.

Reinholt, Mia, Torben Pedersen, and Nicolai J Foss. 2011. ‘Why a Central Network Position Isn’t Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks’. *Source: The Academy of Management Journal* 54 (6). Academy of Management: 1277–97.

Rosen, Sherwin. 1981. ‘The Economics of Superstars’. *The American Economic Review* December: 845–58.

Rosenberg, Nathan. 1990. ‘Why Do Firms Do Basic Research (with Their Own Money)?’ *Research Policy* 19 (2): 165–74. doi:10.1016/0048-7333(90)90046-9.

Rothaermel, Frank T., and Andrew M. Hess. 2007. ‘Building Dynamic Capabilities: Innovation Driven by Individual-, Firm-, and Network-Level Effects’. *Organization Science* 18 (6): 898–921. doi:10.1287/orsc.1070.0291.

Sakakibara, M. 2010. ‘Why Do Some Patents Get Licensed While Other Do Not?’ *Industrial and Corporate Change* 19 (3): 927–45.

Salman, Nader, and Anne Laure Saives. 2005. ‘Indirect Networks: An Intangible Resource for Biotechnology Innovation’. *R and D Management* 35 (2): 203–15. doi:10.1111/j.1467-9310.2005.00383.x.

Sams-Dodd, Frank. 2005. ‘Target-Based Drug Discovery: Is Something Wrong?’ *Drug Discovery Today* 10 (2). Elsevier Current Trends: 139–47. doi:10.1016/S1359-6446(04)03316-1.

———. 2013. ‘Is Poor Research the Cause of the Declining Productivity of the Pharmaceutical Industry? An Industry in Need of a Paradigm Shift’. *Drug Discovery Today* 18: 211–17. doi:10.1016/j.drudis.2012.10.010.

Sapsalis, Elefthérios, Bruno Van Pottelsberghe De La Potterie, and Ran Navon. 2006. ‘Academic versus Industry Patenting: An in-Depth Analysis of What Determines Patent Value’. *Research Policy* 35: 1631–45. doi:10.1016/j.respol.2006.09.014.

Sauermann, Henry, and Michael Roach. 2011. ‘Not All Scientists Pay to Be Scientists : Heterogeneous Preferences for Publishing in Industrial Research’.

———. 2012. ‘Taste for Science , Taste for Commercialization , and Hybrid Scientists’.

Sauermann, Henry, and Paula Stephan. 2013. ‘Conflicting Logics? A Multidimensional View of Industrial and

Academic Science'. *Organization Science* 24 (3): 889–909. doi:10.1287/orsc.1120.0769.

Scherer, Frederic M. 2000. 'The Pharmaceutical Industry'. In *Handbook of Health Economics*, 1B:1297–1336. Elsevier. doi:10.1016/S1574-0064(00)80038-4.

Schmoch, Ulrich. 2008. 'Concept of a Technology Classification for Country Comparisons: Final Report to the World Intellectual Property Organisation (WIPO)'.

Schuhmacher, Alexander, Markus Hinder, and Oliver Gassmann. 2016. *Value Creation in the Pharmaceutical Industry : The Critical Path to Innovation*. John Wiley & Sons, Ltd.

Shah, Neha P, Daniel Z Levin, and Rob Cross. 2018. 'Secondhand Social Capital: Boundary Spanning, Secondhand Closure, and Individual Performance'. doi:10.1016/j.socnet.2017.04.005.

Shepard, A. 1987. 'Licensing to Enhance Demand for New Technologies'. *RAND Journal of Economics* 18 (3): 360–68.

Siciliano, Michael D, Eric W Welch, and Mary K Feeney. 2018. 'Network Exploration and Exploitation: Professional Network Churn and Scientific Production'. *Social Networks* 52: 167–79. doi:10.1016/j.socnet.2017.07.003.

Singh, Jasjit. 2008. 'Distributed R&D, Cross-Regional Knowledge Integration and Quality of Innovative Output'. *Research Policy* 37 (1): 77–96. doi:10.1016/j.respol.2007.09.004.

Song, Yue, Devi R Gnyawali, Manish K Srivastava, and Elham Asgari. 2018. 'In Search of Precision in Absorptive Capacity Research: A Synthesis of the Literature and Consolidation of Findings'. *Journal of Management* 44 (6): 2343–74. doi:10.1177/0149206318773861.

Sorenson, Olav, and Lee Fleming. 2004. 'Science and the Diffusion of Knowledge'. *Research Policy* 33 (10): 1615–34. doi:10.1016/j.respol.2004.09.008.

Sorenson, Olav, Jan W. Rivkin, and Lee Fleming. 2006. 'Complexity, Networks and Knowledge Flow'. *Research Policy* 35 (7): 994–1017. doi:10.1016/j.respol.2006.05.002.

Stern, Scott. 2004. 'Do Scientists Pay to Be Scientists?' *Management Science* 50 (6): 835–53. doi:10.1287/mnsc.1040.0241.

Sternitzke, Christian. 2010. 'Knowledge Sources, Patent Protection, and Commercialization of Pharmaceutical Innovations'. *Research Policy* 39 (6). Elsevier B.V.: 810–21. doi:10.1016/j.respol.2010.03.001.

Stokes, D.E. 1997. *Pasteur's Quadrant*. Washington: The Brookings Institution.

Stuart, Toby E, Salih Zeki Ozdemir, and Waverly W Ding. 2007. 'Vertical Alliance Networks: The Case of University–biotechnology–pharmaceutical Alliance Chains'. *Research Policy* 36: 477–98. doi:10.1016/j.respol.2007.02.016.

Subramanian, Annapoornima M., Kwanghui Lim, and Pek-Hooi Soh. 2013. 'When Birds of a Feather Don't Flock Together: Different Scientists and the Roles They Play in Biotech R&D Alliances'. *Research Policy* 42 (3). Elsevier B.V.: 595–612. doi:10.1016/j.respol.2012.12.002.

Tartari, V., a. Salter, and P. D'Este. 2012. 'Crossing the Rubicon: Exploring the Factors That Shape Academics' Perceptions of the Barriers to Working with Industry'. *Cambridge Journal of Economics* 36 (3): 655–77. doi:10.1093/cje/bes007.

Tartari, Valentina, Markus Perkmann, and Ammon Salter. 2014. 'In Good Company: The Influence of Peers on Industry Engagement by Academic Scientists'. *Research Policy* 43 (7). Elsevier B.V.: 1189–1203. doi:10.1016/j.respol.2014.02.003.

Teece, David J. 2010. 'Business Models , Business Strategy and Innovation'. *Long Range Planning* 43 (2–3). Elsevier Ltd: 172–94. doi:10.1016/j.lrp.2009.07.003.

Thumm, Nikolaus. 2004. 'Strategic Patenting in Biotechnology'. *Technology Analysis & Strategic Management* 16 (4): 529–38. doi:10.1080/0953732042000295829.

Thursby, Jerry G., and Marie C. Thursby. 2001. 'Industry Perspectives on Licensing University Technologies: Sources and Problems'. *Industry and Higher Education* 15 (4): 289–94.

doi:10.5367/000000001101295786.

———. 2011. ‘Faculty Participation in Licensing: Implications for Research’. *Research Policy* 40 (1). Elsevier B.V.: 20–29. doi:10.1016/j.respol.2010.09.014.

Tijssen, R.J.W., Buter R.K., and T.N. Van Leeuwen. 2000. ‘Technological Relevance of Science: An Assessment of Citation Linkages between Patents and Research Papers.’ *Scientometrics* 47 (2). Kluwer Academic Publishers: 389–412. doi:10.1023/A:1005603513439.

Tijssen, Robert J.W. 2001. ‘Global and Domestic Utilization of Industrial Relevant Science: Patent Citation Analysis of Science–technology Interactions and Knowledge Flows’. *Research Policy* 30 (1): 35–54. doi:10.1016/S0048-7333(99)00080-3.

Tijssen, Robert J W. 2010. ‘Discarding the “ Basic Science / Applied Science ” Dichotomy : A Knowledge Utilization Triangle Classification System of Research Journals’ 61 (1972): 1842–52. doi:10.1002/asi.

Torvik, Vetle I., and Neil R. Smalheiser. 2009. ‘Author Name Disambiguation in MEDLINE’. *ACM Transactions on Knowledge Discovery from Data* 3 (3): 1–29. doi:10.1145/1552303.1552304.

Torvik, Vetle I, M. Weeber, D Swanson, and N Smalheiser. 2005. ‘A Probabilistic Similarity Metric for Medline Records: A Model for Author Name Disambiguation.’ *JASIST* 56 (2): 140–58.

Trajtenberg, Manuel. 1990. ‘A Penny for Your Quotes: Patent Citations and the Value of Innovations’. *The RAND Journal of Economics Journal of Economics* 21 (1): 172–87. doi:10.2307/2555502.

Tufts Center for the Study of Drug Development. 2012. ‘Managing the Transition from Nonclinical to Early Clinical Development’. *Tufts CSDD R&D Management Report* 7 (2): 1–4.

Tzabbar, D., and R. R. Kehoe. 2013. ‘Can Opportunity Emerge From Disarray? An Examination of Exploration and Exploitation Following Star Scientist Turnover’. *Journal of Management* 40 (2): 449–82. doi:10.1177/0149206313513613.

Tzabbar, Daniel. 2009. ‘when does scientist recruitment affect technological repositioning?’ *Academy of Management Journal* 52 (5): 873–96.

Wagner, Stefan, and Simon D. Wakeman. 2015. ‘What Do Patent-Based Measures Tell Us About Product Commercialization? Evidence from the Pharmaceutical Industry’. *ESMT Working Paper* 14 (01(R1)): 1091–1102. doi:10.2139/ssrn.2391786.

Wang, Chunlei, Simon Rodan, Mark Fruin, and Xiaoyan Xu. 2014. ‘KNOWLEDGE NETWORKS , COLLABORATION NETWORKS ’,. *Academy of Management Journal* 57 (2): 484–514.

Wang, Jian. 2016a. ‘Knowledge Creation in Collaboration Networks: Effects of Tie Configuration’. *Research Policy* 45: 68–80. doi:10.1016/j.respol.2015.09.003.

———. 2016b. ‘Knowledge Creation in Collaboration Networks: Effects of Tie Configuration’. *Research Policy* 45 (1). Elsevier B.V.: 68–80. doi:10.1016/j.respol.2015.09.003.

Waring, Michael J., John Arrowsmith, Andrew R. Leach, Paul D. Leeson, Sam Mandrell, Robert M. Owen, Garry Pairaudeau, et al. 2015. ‘An Analysis of the Attrition of Drug Candidates from Four Major Pharmaceutical Companies’. *Nature Reviews Drug Discovery* 14 (7). doi:10.1038/nrd4609.

Williams, Richard. 2017. ‘Analyzing Rare Events with Logistic Regression’.

Woolf, Steven H. 2008. ‘The Meaning of Translational Research and Why It Matters’. *JAMA* 299 (2). American Medical Association: 211–13. doi:10.1001/jama.2007.26.

World Health Organization. 2010. ‘How to Conduct Patnet Searches for Medicines, a Step-by-Step Guide’.

Zahringer, Kenneth, Christos Kolympiris, and Nicholas Kalaitzandonakes. 2017. ‘Academic Knowledge Quality Differentials and the Quality of Firm Innovation’. *Industrial and Corpo* 26 (5): 821–44. doi:10.1093/icc/dtw050.

Zucker, L G, and M R Darby. 1996. ‘Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry.’ *Proceedings of the National Academy of Sciences of the United States of America* 93 (23): 12709–16.

Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong. 1998. 'Geographically Localized Knowledge Spillovers or Markets' XXXVI (January): 65–86.

———. 2002a. 'Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology'. *Management Science* 48 (1): 138–53. doi:10.1287/mnsc.48.1.138.14274.

———. 2002b. 'Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology'. *Management Science* 48 (1): 138–53. doi:10.1287/mnsc.48.1.138.14274.

Zucker, Lynne G, and Michael R Darby. 1996. 'Costly Information: Firm Transformation, Exit, or Persistent Failure'. *The American Behavioral Scientist* 39 (8). ProQuest Central pg: 959.

Zucker, Lynne G, Michael R Darby, and Jeff S Armstrong. 2002. 'Commercializing Knowledge : University Science , Knowledge Capture , and Firm Performance in Biotechnology', no. October 2014.

Zuniga, M., and D. Guellec. 2008. 'Survey on Patent Licensing: Initial Results from Europe and Japan'. *Internal Working Document OECD, Paris*.

Doctoral dissertations from the Faculty of Economics and Business, see:
http://www.kuleuven.ac.be/doctoraatsverdediging/archief.htm