UHASSELT

KNOWLEDGE IN ACTION

Faculty of Business Economics

Master of Management

Master's thesis

Open innovation: which role remains for firms research department?

Steuwie Fonke

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Strategy and Innovation Management

SUPERVISOR :
dr. Relinde COLEN

2020
2021

UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT

KNOWLEDGE IN ACTION

Faculty of Business Economics

Master of Management

Master's thesis

Open innovation: which role remains for firms research department?

Steuwie Fonke

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Strategy and Innovation Management

SUPERVISOR :

dr. Relinde COLEN

Disclaimer

This master thesis was written during the COVID-19 crisis in 2020. This global health crisis might have had an impact on the (writing) process, the research activities and the research results that are at the basis of this thesis.

Preface

This report was written in light of my master dissertation for the Master of Management at the University of Hasselt. The topic, allocated to me by the university is "Open innovation: which role remains for firms' research department?". With this dissertation I hope to highlight the key elements around the research topic and contribute to the knowledge around it.

This dissertation has come with its fair share of hurdles, not the least of which the global pandemic that made it impossible to meet with any one of the faculty in person. However I feel that I have learned a lot during the writing process, some examples being the statistical analyses needed for this thesis and the very interesting literature around innovation in the pharmaceutical industry.

I would like to express my graduate to my promotor, dr. Linde Colen, for her guidance during this research. I am convinced that without her feedback and encouragement on this thesis, and especially her help during the empirical part, this thesis would not have come to be.

Summary

In recent years the role of the research department within pharmaceutical industry is changing. Since the introduction of the open innovation paradigm firms are increasingly sourcing research and innovation trough external ways (Chesbrough, 2003). All the while investments in research itself by pharmaceutical companies is decreasing (Arora, Belenzon and Patacconi, 2015). This might have some very profound effects as investments in in-house research capabilities leads, inter alia, to absorptive capacity (Cohen and Levinthal, 1989; Cohen and Levinthal, 1990). This dissertation thus probes into the independent role remaining for internal research departments and aims to answer the question : "For which research do internal research departments independently contribute to the publication stock within drug discovery?".

In order to answer this question I conducted an explorative research, exploring the differences utilizing some descriptive analyses. The non-directional hypothesis extends on the research question, stating: There is a difference between the characteristics of independent and collaborative publications across the pharmaceutical industry. To test this, *two group t tests* were used to test the significance of the mean difference of the various characteristics relating to independent and collaborative publications. To strengthen our findings three regressions analyses are included in the dissertation. The dataset used consists of 59023 publications in pharmaceutical journals published between 1994 to 2002 by 62 companies. These companies are the largest R&D spenders in the pharmaceutical industry from the 2004 EU scoreboard.

The first difference I found is that on average the independent publications tend to be more basic compared to collaborative publications. Moreover data shows that the likelihood of a publication being independent diminishes if the publication is more recent or when the number of authors is higher. Independent publications on average receive less citations from other scientific publications, suggesting that the independent publications have less scientific quality. This difference is even more apparent within independent publications as a category. I find that basic independent publications receive significantly more citations then independent applied publications.

The analyses suggests that relative to all citations by patents independent publications get cited less by their own publishing firm. Indicating that in fact firms don't have any heightened interest in their own publications be it technically or commercially. Moreover I find that generally it seems that independent publications are referred less to by patens. This also goes for clinical projects and firm clinical projects. With clinical projects again I find that relative to all publication the independent publication get cited less on clinical projects.

When looking at the attributes of the patents referring to the publication we also find some differences when comparing independent and collaborative publications. On average the patents referring to independent publications in the dataset report a higher average number of citations of other patents. This seems especially true for basic independent research publications. These results however do not carry over to our regressions, as we do not find any significant effect of a publication being independent on the citation of the patents referring to the independent publication. What we do however conclude is that the knowledge sources of patents referring to independent publications on average is lower than in those patents referring to collaborative

publications. The average maximum citations of patents referring to independent publications is also higher as opposed to collaborative publications. Lastly we find that the average time between publication and citation is higher for independent publications.

These analyses thus support the hypothesis. Concluding that there is a difference between the characteristics of independent and collaborative publications across the pharmaceutical industry. I see three reasons that could possibly explain these findings, and explain why there is a difference in the characteristics of independent and collaborative publications. These three reasons are however not necessarily mutually exclusive.

The first explanation I propose is that these differences are cause by the fact that firms deploy their internal research department to fill in some very specific gaps in the literature in their own strategically relevant core research field. Moreover the differences could be due to the search of firms toward disruptive and/or breakthrough innovations. The literature review shows that as science advances it is getting harder to innovate (Jones, 2009) furthermore firms have an increasingly narrower scope (Arora, Belenzon and Patacconi, 2015; Jones, 2009). This could explain why the independent publications are more often basic research publications, as this is fundamental in drug discovery. Furthermore this would also explain the lower citations and forward development of the publications, as these publications might not be very necessarily very interesting for other firms or institutions.

Another potential explanation for these findings could be the influence of (star) scientist within the internal research department and their own inclination for basic research and certain research domains. In the literature I explore the influence of researchers on the publication stock of private firms, already research by many scholars. Allowing scientist to publish is an argument to convince these research to work within private firms (Henderson and Cockburn, 1998). These researchers have very specific fields of expertise (Jaffe, 1986). This could be an explanation for the somewhat strange finding that firms do in fact not cite or develop their independent publications more in comparison to collaborative publications. I find that firms in fact develop their own publication less in to clinical projects and even patents when compared to all publications.

Absorptive capacity plays a very important part role in private scientific research, and the differences we observe could also be a result of the effects of absorptive capacity. I argue that the differences found are a results of firms trying to sustain or improve their absorptive capacity. As the literature shows firms are increasingly relying on absorptive capacity to fuel their innovations as open innovation emerges (Chesbrough, 2003). This consideration is further supported by the relationship found in the regression analyses between publications being independent and the investments of a firm and the total amount of publications and clinical projects of the firm relative to tehri R&D investments.

The findings I present are limited to the pharmaceutical industry, and the dataset only amounts for the 63 largest spenders R&D wise in Asia, Europe and North-America between 1994 and 2002. This of course constrains the generalizability of the findings. Furthermore the methodology of this study also presents some limitations as students t test is in some cases a less then optimal statistical test.

The results serve not as conclusive evidence but present a broader direction for further research. This study as is set up as an explorative one, due to a lack of prior research towards the role of internal independent research. Further research could direct its attention towards testing the reasonings we propose for the differences. Furthermore there could be still more differences and inter related effects when comparing independent and collaborative research, so research could direct its attention on other characteristics of publications. Moreover quantitative research could uncover more interrelated effects of independent publications, for example on further development of publications.

Table of contents

List of tables

List of graphs

Introduction

Since the general paradigm of innovation is shifting more and more to open innovation firms are relying increasingly on external partners and ways to source innovation and its corresponding research (Chesbrough, 2003). Research has by many scholars been confirmed to be the source of technological advancement and accordingly economic value for the firms that produce it (Griliches, 1998; Cockburn and Henderson, 1998) and even society at large (Nelson, 1959). However in modern times we see a decline in the investment towards research by private firms (Arora, Belenzon and Patacconi, 2015). This trend might have many implications for industries relying on research such as the pharmaceutical industry. One element that is for example very important in private research is the absorptive capacity generated by having an internal research department (Cohen and Levinthal, 1989; Cohen and Levinthal, 1990).

This paradigm shift and the decline of interest and investments towards research by firms leaves us to wonder what role is there still for the internal research departments. To get a view on this, this dissertation aims to answer the question: "For which research do internal research departments independently contribute to the publication stock within drug discovery?".

This study will focus on the publications of pharmaceutical firms in scientifical journals. Readers will find that the literature study and all results are thus limited or tailored to the pharmaceutical industry. To answer the research question a quantitative approach will be used, using data on publications from Asian, European and North-American pharmaceutical firms in scientific journals. The dataset includes information on references towards the publications and forwards citations. The goal is to find differences in publications that are made independently by the firm and those publications that are realized in collaboration. With these findings the dissertation provides an explorative argumentation of the differences and their impact. Furthermore I will provide some suggestions for further research on this topic.

Firstly there is a literature study, further underlining the importance of this topic and the relevance of it for the pharmaceutical industry. In this first section I will also make some important distinctions regarding research. In the second part of the literature study I will briefly go over some rationalities retaining firms form research. After this follow the reasons for firms to actually still adopt private research despite the reasons not to invest. Lasty the literature study gives insight on the existing literature regarding internal research and research investment throughout recent years and the emerging open innovation paradigm. This is followed by the empirical part, which was briefly discussed already. The dissertation will end with a conclusion, linking the findings to existing knowledge and giving an explorative reasoning as to what is the independent contribution of internal research departments to the publication stock within drug discovery.

Literature study

Innovation and science in the pharmaceutical industry

Before we look further in to research within the pharmaceutical industry we need to establish the relationship between innovation and science itself, as this is an important one. This link is why entities, for profit or non-profit, private or public, invest in research (Salter and Martin, 2001). This important relationship however is not an easy one to make because as said by Jaffe in his 1989 paper "knowledge is, after all, a public good". In this same paper Jaffe (1989) also explores the effects of university proxy on local patenting behaviour, finding a significant effect of university research on corporate patents. The link between research and innovation is also more broadly found in the relationship between accumulated academic science and productivity within industries (Adams, 1990; Mansfield, 1980). However this productivity increase thanks to academic science does experience a time lag of about 10 to 30 years before it transforms in to productivity. The question then is how does the effect come to be. Many scholars attribute this to knowledge spillovers (Salter and Martin, 2001).

While talking about how innovation advances and grows some scholars wonder as to where the limit of these phenomena and their effect on firms and more broader the economy lies (Jones, 2005; Gordon, 2012). This is of interest as firms would withdraw from science if it's effect on innovation would diminish (Gordon, 2012; Jones, 2009). However the literature argues that although the advancement of innovation fuelled by science will not vanish, there are some factors that negatively influence the pace of these advancements. Gordon (2012) for example identifies 6 of these 'headwind' factors. He identified firstly the fact that hours worked per capita are decreasing as more and more baby-boomers retire and life expectancy is getting prolonged in comparison to the retirement age. This results in a slower growing output per capita and by default must grow slower compared to productivity. A second influence is the plateau of schooling grade caused by costly schooling and rapid growth in previous decades. Moreover, innovations require knowledge generated by people that have undertaken higher education (Jones, 2009). The third factor that Gordon reports is the growing inequality. He states that if at all we care about consumer wellbeing we must also look at those with the lowest incomes. The growing discrepancy however, Gordon argues, has negative effect on the average growth rate of GDP per capita. The fourth factor is globalization, with special attention to its interaction with IT. In an ever more connected world it is all too easy to outsource or automate jobs. This will have a damaging effect on dominant economic countries that have the highest wages. The second to last factor mentioned in the paper is energy and het environment. As scientists urge us to keep the effects of pollution and other climate changing factors in mind this nudges firms to rethink their business processes and invest in a different way than ever before. Lastly Gordon (2012) identifies the increase of debts within private households as an increasing problem, as an unprecedent stress on the economic system. According to Gordon (2012) his findings are mostly limited to the United States and states that although there is a possibility that these factors play a role outside of the US, a generalization is not within the scope of his research.

Next to the relationship between innovation and science itself we also found that scholars often make a distinction between two research types (e.g. Rosbeberg, 1990; Mansfield, 1980; Leten, Kelchtermans and Belderbos, 2010; Lim, 2004). The distinction is between basic and applied research. As with many definitions within scientific literature the definition of basic research is not one that is written in stone. There is abundance of definitions for basic research (Rosenberg, 1990). Multiple studies cited the National Science Foundation for their definition of basic research. Their definition as: "A *systematic study directed towards greater knowledge or understanding of the fundamental aspects of phenomena and observable facts without specific immediate commercial applications in mind, although research may be in fields of present or potential commercial interest of those performing the research activities* " (Mansfield, 1980; Leten, Kelchtermans and Belderbos, 2010; Lim, 2004).

The distinction between basic and applied research is not as obvious as it seems, as it seen as something operationally (Reagan, 1967). When comparing the insights of the distinction made by Raegan (1967) with the NSF definition (Mansfield, 1980) it is mostly the lack of commercial application that distinguishes applied research from basic research. Furthermore Mansfield (1980) clearly indicates a difference between applied and basic research when it comes to expenditure. Mansfield (1980) that the budget composition for R&D is changing for almost all industries, with firms cutting the proportion of their budget going towards basic research. Furthermore Mansfield finds evidence indicating that the combination of applied and basic research has benefits. Lastly his finding indicate that basic research relates to some extent to long term R&D.

Within the pharmaceutical sector there is a very well described R&D flow, with a clear distinction between basic and applied research (Ward & Dranove, 1995). Ward and Dranove discuss the research and development chain in the pharmaceutical industry in their 1995 paper. They state that most basic research is conducted in early stages of drug development, resulting in new chemical or biological compounds which are not necessarily tailored towards a specific drug or treatment. If these findings result from public basic research they are published in the scientific community, then firms can obtain the rights of these compounds. This brings the firms to the development stage, where applied research will be conducted. Entailing that firms search for a therapeutic use for the compound or substance. This, contrary to basic research, is tailored towards a disease. Here the firm can apply for a "use" patent and start clinical trials, this has to be up to standard of the regulators of a specific country. In the USA these guidelines are set and controlled by the FDA, which can grant permission to bring a drug to market.

Insert figure 1 here

The R&D flow of pharmaceuticals depicted by Kaitin (2010) is more extensive than the flow described by Ward and Dranove (1995). The flow of Kaitin (2010), shown in figure 1, includes the effects of outside information by academia, small pharmaceutical and biotechnology and other partners within a fully integrated pharmaceutical network model of drug development (Kaitin, 2010). The flow indicates the different research stages in the green arrows and builds upon this with the roles of the outside influences during these stages.

For scientific research to be studied within companies scholars need a measurement. Often these scholars look towards patenting behaviour of firms and the research publications of pharmaceutical companies in scientific journals. However with both these measures in mind the difficulty of tracing science remains (Arora, Belenzon and Patacconi, 2015). Research publications is considered a measure of investments of basic research, however it could be seen as a mere reflection of publication behaviour (Arora, Belenzon and Patacconi, 2015). Publication behaviour of a firm describes the tendency of firms to publish their findings in a scientific journal. One assumption that is made by Arora, Belenzon and Patacconi is that publication of articles is depended on the information that it contains. They state that patentability and the type of research influences the publication. In their view commercially sensitive information tents to be published less.

Patents are found to mirror basic research intensity of a firm, especially in the pharmaceutical sector (Lim, 2004). With patent protection firms can protect their intellectual property, but it is inaccurate to state that this thus directly mirrors the innovation performance (Griliches, 1990). Lim (2004) finds that firms who conduct more basic research tend to cite more basic scientific knowledge, and suggest that this has an impact on absorptive capacity. The concept of absorptive capacity will be discussed more in the next section.

Disinclination of companies to invest in research

Historically seen most basic research was conducted in non-profit laboratories (Nelson, 1959). In 1953 roughly sixty percent of all basic research was conducted by non-profit institutions according to Nelson. Corporate investments in to basic research was lagging behind, corporations were not willing to invest in basic research. However Nelson (1959) disclosed the societal benefits of basic research, and stated that scientific research could potentially lead to economic value. In his research he also looked at the factors withholding companies from conducting basic research at the time. The factors that Nelson has identified have also been confirmed in other research, and in fact has been recognized as the main explanations for the fact that research was mostly a publicly funded case (Pavitt, 1991).

Nelson (1959) states that companies averted from this practice because of the unpredictability of research and the cost that it presented. Conducting research presents companies with a lot of unpredictability and uninsurable risk (Rosenberg, 1990; Arrow, 1962). Firms allocate a lot of resources towards their innovation process with no guarantee that this will yield a result, this is a real risk and discourages companies to undertake these innovation processes (Arrow, 1962). In the pharmaceutical industry. The extensive testing and difficult approval process of drug regulators results in an average of just 16% of compounds discovered, and that are subsequently submitted for approval, to reach the market. According to Kaitin (2010) is a decline from the 21.5% approval rate in the 90's. This long development time and the low level of compounds actually making it to market logically means a high overall research cost. This entails an escalation of R&D expenditure within the pharmaceutical market since the 80's. The cost of developing and bringing a new conventional pharmaceutical product to market is 1.32 billion US dollars (Kaitin, 2010). So it is essential for competitiveness within pharmaceutical industry to reduce the time and resources it takes to develop novel drugs and compounds.

Furthermore there is a time lag between the significant investments of basic research and its pay-offs (Nelson, 1959) . Especially within the pharmaceutical industry investments take a long time before they pay of (Rosenberg, 1990). There is a long average duration for drugs to be developed, clinically tested and approved. On average it takes nearly 9 years from the start of human testing to market. (Kaitin, 2010).

Nelson (1959) also stated that firms feared that their investments in to basic research would be exploited by other firms. There is the risk of so called "free riders" exploiting the knowledge (Rosenberg, 1990; Arrow, 1962). This is a simple as it sounds, other firms taking the knowledge that the firm generates and appropriating it within their own company. This can take place through a number of mechanisms such as publication but also mobility of personal can increase this risk (Rosenberg, 1990; Arrow, 1962.

Advantages of in-house research

Despite these factors unmanning firms from conducting basic research it has been on the rise throughout the '60s '70s and 80's (Pavitt, 1991). So there must be reasons for firms to take on research activities. The following section of the literature study will go in depth on reasons for firms to conduct research, and the advantages it can bring to firms in the pharmaceutical industry.

Like research within business itself the literature regarding basic research within business was also on the rise throughout the '60s '70s and 80's. Including implications of basic research, both private and public, on public policy. Arrow (1962) stated for example that it is in the public interest for new knowledge to be available free of charge. In his view this would be beneficial to the general economy, however it would eliminate all incentive for business to innovate.

This is in line with the previous conclusion drawn by Nelson (1959), however research can be incentivized by the option to protect the products of research under intellectual property right. With this property right firms can legally ensure that the information can only be exploited within the firm, and thus work towards creating benefit for the firm (Arrow, 1962). However this leads to the underutilization of that knowledge, as other firms cannot benefit from the knowledge. This has negative ramifications for both business and society. This phenomenon is referred to as the appropriating problem. However not all knowledge can be protected legally protected leading to so called spill overs, where firms that did not participate or co-fund any privately funded research still exploit the knowledge generated by it (Rosenberg, 1990).

Nelson argues, as mentioned in the previous section, that science has some fundamental societal benefits. According to his viewpoint most social problems would be solved if everyone would take a scientific point of view for every activity. Furthermore Nelson states that science is what allows use a s a society to produce value and to research the creation of economic. He links the value of the output flow in economy to science.

We have to look at basic research as an investment for companies, as they won't spend money for public-spirited reasons (Rosenberg, 1990). Furthermore Rosenberg (1990) states that while firms will likely not be able to exploit all potential benefits of the results of their research, the ability to

capture even some of the benefits should suffice. If the potential results of invention exceed the uncertainty of invention and research, firms should pursue it (Arrow, 1962). Even with possible knowledge spillovers in mind the potential business opportunity can yield a huge return on investment (Rosenberg, 1990).

In his paper Rosenberg (1990) highlights two factors driving basic research in companies. Firstly he states that the part of the knowledge or research is unintentional. The research is an unintentional by-product of research towards a specific purpose. Secondly firms must conduct basic research in order to provide them valuable direction for the evaluation and exploitation of outside knowledge. Basic research is an essential tool to utilize external science and tap in to the public knowledge that is generated by academia and non-profit institutions.

This last assumption is also supported in case studies of large US pharmaceutical firms (Gambardella, 1992). These case studies stated that firms with a strong in-house scientific capabilities are better in exploiting external science. Merck for example acknowledge the benefits that external can have on innovation, and organized their research departments like the scientific community (Gambardella, 1992). This positive effect of in-house scientific capability has been linked to absorptive capacity (Cohen and Levinthal, 1989; Cohen and Levinthal, 1990). Absorptive capacity is the capacity of a firm to assimilate external knowledge, it's effectiveness of exploiting this knowledge and its ability to match other innovations or knowledge (Cohen and Levinthal, 1989; Cohen and Levinthal, 1990). This is supported by other scholars as well stating that if firms correctly develop their absorptive capacity it will result in a guided technical search and relevant external technology (Arora, Belenzon and Patacconi, 2015). In their paper Cohen and Levinthal (1989) argue that internal research within business has two major outcomes. One of these outcomes is the positive effect of internal research on absorptive capacity. They see this as the two faces of R&D, as also stated in the title of their publication. They state that these two faces of R&D will have a varying effect on the incentive of firms to invest in research depending on the complexity of their industry or company and on other attributes within the firm. Furthermore their research proved that the absorptive capacity of a firm has a moderating effect on technological opportunity and appropriability. Moreover its concluded that absorptive capacity has an important mediating effect on external knowledge assimilation. Firms are suggested to be impacted by the learning environment in which they operate, making absorptivity capacity a relevant factor in the decision of allocating resources (Cohen and Levinthal, 1990). For example innovations that build upon pre-existing information are observed to be adopted quicker possibly by the ease of learning (Cohen and Levinthal, 1990).

Great absorptive capacity can lead to two advantages for firms. Firstly it could lead to a first mover advantage, as firms with great internal research capacity can exploit the findings of academia and public laboratories (Cohen and Levinthal, 1989). This is because those firms can, trough the mediating effect described previously, more easily identify recognize, assimilate and commercialize this knowledge. Furthermore Cohen and Levinthal (1989) state that a greater internal research capacity can lead to a swifter second mover response. Spill overs from other companies and even competitors can be facilitated within the company to quickly follow the innovation at hand. However in a other case study of pharmaceutical companies it is argued that investing in in-house

basic research alone is not sufficient to access all of these advantages). Arora, Belenzon and Patacconi (2015) state the pay offs off absorptive capacity only emerge if there is a sustained investment in to internal research, and even then a time lag between investments and pay-off remains.

In a more recent study absorptive capacity has been reconceptualized and focuses on the gain and sustenance of a competitive advantage through the ability to create and exploit knowledge (Zahra and George, 2002). Additionally, the same study states that absorptive capacity should be seen as a dynamic capability consistent of four dimensions. The distinction between a capability and a dynamic capability being the focus on organizational change. The four dimensions defined by Zahra and George (2002) are acquisition, assimilation, transformation and exploitation. In their view these four dimensions build further upon each other within the firm. Their model differentiates between potential and realized absorptive capacity, stressing the importance of the firms capability to transform and exploit their acquired and assimilated knowledge in order for it to generate profit.

Cockburn and Hendersons 1998 study argues that for a firm to use external knowledge towards improved research productivity it must also actively collaborate with the public sector and it's research departments. The research finds that the quality of researchers within a private company is of great importance for the access to public sector research and the quality of internal research. This quality is expressed on the basis of the standing in research hierarchy, and their connecting and engagement with public sector research. Moreover Arora, Belenzon and Patacconi (2015) also state that greater absorptive capacity might lead to increased attractiveness of the firms towards scientist.

This last element in itself can be a another reasons for firms to perform basic research. Not only do scientist (especially so called 'star scientists', the best of their field) contribute to innovation through their university work they also contribute to innovation via collaboration (Zucker et al., 1998). So Zucker suggest that by collaborating with these scientist pharmaceutical companies increase their innovation performance. This is due to the capabilities of these experienced scientist (Jaffe, 1986). Firms can attract these scientist by allowing them to publish their findings (Henderson and Cockburn, 1998). This is valued by these scientist as it allows them to build a ranking in the academic world, furthermore these scientist might have an inclination to favour basic research and by allowing them to pursue this inclination it becomes more attractive for them to work in the private pharmaceutical sector (Henderson and Cockburn, 1998).

Another possible reason to conduct research is its positive impact on productivity growth and profitability (Griliches, 1998). According to Griliches (1998) data shows a positive link between productivity growth and profitability. Griliches states that this is especially true for privately financed R&D, in comparison to publicly financed R&D efforts on the firm level. These findings are also supported by Arora, Belenzon and Patacconi (2015), who explain the general logic behind this effect. Arora, Belenzon and Patacconi (2015) state that scientific research drives technical advancements, and that without research scientific and technical advancement would break off. Firms need these advancements, in the form of publications, in order to fuel their own innovations. In a study regarding SME's it was found that basic research can lead to competitive advantage for firms, by creating differentiation from competitors, customer loyalty and product innovation

(Rosenbusch, Brinckmann and Bausch, 2011). This research also showed that innovation can help reduce entry barriers in to otherwise difficult industries.

Internal research and research investments throughout time

The following section sets out to describe the evolution of internal research and research investments throughout time, to give more insight in to the broader field of interest of this paper. This evolution throughout time is important to understand firstly the current research expenditure of pharmaceutical firms, this information is relevant as it will also be carried through to the analysis of the empirical part of this research. Furthermore I will describe a paradigm shift in the innovation landscape that has had a major impact on the internal research department and its function within the research funnel.

The understanding in the 90's was that corporate basic research by firms was performed within a small section of firms (Rosenberg, 1990). Furthermore the same research indicated that this basic research was also concentrated mostly in a small number of sectors. These sectors included chemicals, electrical equipment, aircraft & missiles and machinery, however is should be noted that the data used to derive these sectors was incomplete. However as previously mentioned early researched showed few firms actually conducting basic research (Nelson, 1959) but this has been rising all trough '60s '70s and 80's (Pavitt, 1991). In this next section I shall thus look in to the spread of research within private firms, with particular attention to its spread in the pharmaceutical sector.

This literature review will look at the evolution of investment towards research within private companies only after the second world war, this is also the case in many other reviews and papers. This mainly caused due to a lack of data prior to this period, as the NSF was not established until 1950's (Mowery, 2009) and the lack of interindustry heterogeneity prior to 1940 (Lee, 2003). After these years of conflict (1939-1945) the biomedical industry among others started to grow at a very high rate, moreover trade and knowledge flows got increasingly more global (Mowery, 2009). A key characteristic of R&D in this period, as identified earlier in this literature review, is the fact that is was predominantly government funded (Nelson, 1959; Mowery, 2009). This funding however was mostly used towards universities and industry focused research as opposed to government laboratories, with a significant focus on defence research (Mowery, 2009). However Mowery (2009) also states that R&D by industry kept on growing in comparison to federal R&D, and has even surpassed the federal R&D budget.

Mowery (2009) found that post world war II a shift around knowledge and industry was happening, instead of Western-Europe being the focal point this was now shifting towards North-America. This same trend is identified by Hagedoorn (2002), who also find that since 1960 companies are increasingly developing R&D partnerships. Together with the IT sector the pharmaceutical sector is at the forefront of this trend. The study finds that R&D partnerships are mostly between large firms, the paramount of which are located in the US, and other smaller countries (Hagedoorn, 2002). This trend in in line with the findings of Mowery (2009), who stated that an ongoing internationalization was accruing. Evidence suggest that during the 1960's companies used R&D as a competitive advantage within the pharmaceutical industry, for example

innovators that came up with antibiotics continued to increase their R&D intensity after successfully brining their product to market (Lee, 2003).

During the 1970's R&D and especially basic research had a strong impact on the productivity growth of private firms (Griliches, 1998). Meanwhile data from the National Science Foundation on that same decade showed that R&D expenditure peaked in 1968 and showed an overall slowdown of R&D expenditure growth (Griliches, 1998). On basic private research the NSF data even showed a decline of expenditure, which Griliches points out to be costly when combined with the knowledge that basic research had a strong positive effect on productivity growth. It must be noted however that during this period the global economy, and the US economy, experience a prolonged recession.

Arora, Belenzon and Patacconi (2015) conducted a detailed study on corporate research over the period 1980-2007. Their first finding was that the scientific publications by firms was steadily declining. Moreover the data in their study shows a drop in basic research investments within large American firms when comparing the figures of 1980 to those of 2007. However these same firms don't show an overall decrease in their R&D investments. They conclude that the willingness of established companies to conduct research is decreasing. Arora, Belenzon and Patacconi state that this shift is due to the fact that firms value patents and the results of science but neglect the actual scientific capabilities. They attribute this to globalization and firm scope as opposed to publication tendency and the perceived value of science as a driver for innovation.

The decrease in the willingness of companies to invest in research manifests itself, among other factors, in the propensity to publish (Arora, Belenzon and Patacconi, 2015). The results in their study state that overall firms are publishing less, especially when looking at basic research. With these findings they state that the fact that the decline in publication cannot be explained by a change in publication behavior. Instead it shows that large firms are actually changing the composition of their research, conducting less basic research and emphasizing more applied and patentable research. The data in the study of Arora, Belenzon and Patacconi (2015) shows that decline of publishing is steeper in the pharmaceutical sector.

Insert figure 2 here

As already described patens are an indicator of innovation output. However, and this is also stated by Arora, Belenzon and Patacconi (2015), is a very limited indicator. In their study patents are used as an approximation for the application of science towards commercial usage. They did not see any major shift of patent citations to science over time. When looking at patens they give special importance to the age of citations within the patents. In their view scientific capability enables firms to use more recent studies. Furthermore they link this to publishing firms, stating that these firms are more likely to cite recent publications. However Arora, Belenzon and Patacconi suggest that it is possible that pattens are citing older publications, implying that firms might decrease investment in research if the knowledge needed is older. Furthermore they imply that if a firm is able to access external knowledge through other means then firms might be inclined to further reduce internal research investments. However the research of Arora, Belenzon and

Patacconi (2015) states that they find no evidence of decreasing absorptive capacity, or that within the scope of patents science has become less relevant.

Another factors influencing the rate of innovation found within the literature is the principle of 'burden of knowledge' (Jones, 2009). Jones (2009) states that because of previous innovations and scientific advancements it is getting harder to innovate. He states this concepts exhibits itself within firms by the fact that teamwork is occurring more and more and the age of individuals that come up with an innovation for the first time is rising. Another indication might be the fact that people are becoming more specialized (Jones, 2009).

Arora, Belenzon and Patacconi (2015) attribute this withdrawal from science to globalization, this is in line with one of the six factors identified in the literature diminishing economic growth (Gordon,2012). Their data indicates for example the significant reduction of R&D expenditure within large firms in sectors with an increased Chinese import. Meanwhile these same firms show an increased propensity to patent and a decline in the stock market value of their publications. Their study states that globalization might have two major impact in this regard. Firstly the increase of competition that globalization brings might put pressure on firms, who in turn react by reducing their investments in research. This is a very short termed vision, but firms might do this if they perceive a decrease of value of innovation. Another factor could be that the increased competition leads to an overall decrease in funds within the firm, resulting in decreased funds available for research. These implications however are only observed in association and might possibly not have causal reasons. Arora, Belenzon and Patacconi (2015) state that it was not the objective of their study, rather they wanted to explore the mechanisms at hand.

The second factor attributed to the withdrawal from science to is firm scope (Arora, 2015; Jones, 2009). Due to this growth firms are increasingly focusing on their core business. This includes the adoption of a narrower product portfolio and taking on a smaller part of the value chain. Whilst this might have some advantages it also makes basic research and its unpredictable and broad outcome less valuable for firms. They argue that this trend has been manifesting ever since the 90's and that overall les diversified firms derive less value from their scientific capability and reduce their investments in science in line with their scope. The research of Arora, Belenzon and Patacconi (2015) also found that firms with a smaller scope also tend to publish less in journals, however this same trend is not present when looking at patents, at least not a statistically significant level. Similarly the stock market value of publications is related to a firms scope over time, but this does not hold for the stock market value of patens.

The study on the decline of willingness within large firms to invest in to scientific capability concludes with pointing out that this decline is reflected in the propensity to publish but also the implications for acquisitions and mergers. Arora, Belenzon and Patacconi (2015) state that their findings, in line with other literature (e.g. Arora and Gambardella, 1990; Higgins and Rodriguez, 2005; Mowery, 2009), indicate that firms increasingly exploit and assimilate external knowledge through alliances and licensing. In their 1990 paper for example Arora and Gambardella suggest that whilst previously innovations where made by in-house R&D departments this has shifted between the 70's and 90's towards innovations being produces outside of the company. Moreover they suggest four ways firms realize this external research, especially within pharmaceutical and

biotechnological research. The first strategy being research and development in conjuncture with other companies, secondly research in cooperation with universities followed by venture capitalism and acquisitions respectively. Moreover these four strategies are increasingly shifting R&D efforts of an industry from bigger to smaller companies, increasing the networks within industries and strategic alliances created by firms (Mowery, 2009) . There are significant differences between these four strategies regarding objective and focus (Arora and Gambardella, 1990). Their papers concludes that these strategies all fulfil a distinct function in external innovation and must been seen as complements to one another. Another studies show that in pharmaceutical companies showing a decline of internal research productivity acquisitions are effective at realizing positive returns (Higgins and Rodriguez, 2005). This same paper also concurs with Arora and Gambardella on the fact that these acquisitions are a compliment to other R&D efforts, especially the internal R&D department. They conclude with the fact that the negotiations and prior access to information regarding the R&D within the firms that is acquired is the key to success with this outsourcing type (Higgins and Rodriguez, 2005).

If research is not taking place outside of the company as substitute for internal research the decline of investments towards scientific capability could indicate that private research itself is in decline. And this arguably is a more worrisome scenario, as the knowledge generated by these firms play a part in the scientific and technological advancements fuelling firm and economic growth (Arora, Belenzon and Patacconi, 2015). Moreover they state that firms likely compensate this decline by other means such as acquiring smaller firms our strategically forging an alliance with a university. This indicates a reallocation of resources within the firm and more generally on industry level. Specialized organization and effective smaller firms are forming a source of corporate research. This trend is also indicated by a great amount of papers regarding open innovation (i.e. Chesbrough, 2007) which I will discuss in the next section.

I have discussed already briefly the increasing amount of pharmaceutical firms looking at external ways of assimilating knowledge identified in the literature (e.g. Arora, Belenzon and Patacconi, 2015: Arora and Gambardella, 1990; Higgins and Rodriguez, 2005; Mowery, 2009). This trend is a key element of what is called Open Innovation, a term first coined by Henry William Chesbrough in his book 'Open Innovation, The New Imperative for Creating Profiting from Technology' (Chesbrough and Bogers, 2014). Open Innovation can be introduced by the example of Lucent (an AT&T spin-off) and Cisco, active in the telecommunications market (Chesbrough, 2003). Chesbrough explains in his book how these two companies both did very different things when it came to research within the firm. Lucent heavily relied on their internal R&D department in order to innovation whilst Cisco, at that time still a starting company, did not operate large scale internal research. Regardless Cisco managed to keep up with the innovations of Lucent and at times was even a step ahead of them. Cisco would invest, collaborate or even acquire other start-ups for their innovations. This way they kept up to speed with innovations in the industry, without investing in the research of its own.

The following will discuss Open Innovation and its emergence, along with key characteristics of both Open and Closed Innovation. The main purpose of this section is to highlight some key

characteristics of the ecosystem in which the pharmaceutical have to operate, especially their internal research department as focus of this research.

Let's firstly explore Closed Innovation, the traditional innovation model adopted by most companies before the shift towards Open Innovation (Chesbrough, 2003). According to Henry Chesbrough (2003) the purpose of closed innovation can be summed up by the virtuous circle: firms make a fundamental technology breakthrough, implement this in new products and features, consequently realize an increase in sales and profit allowing them to increase investments in R&D leading to even more fundamental technology breakthroughs. These breakthroughs then trigger the same sequence all over again.

Chesbrough (2003) identifies 4 different erosion factors, encouraging firms to step away from closed innovation. The first erosion factor identified by Chesbrough (2003) is the increasing availability and mobility of skilled workers. The liberalization of higher education has provided the labor market with more highly skilled workers and post graduates. Employees change jobs more often, this results in diffusion of their knowledge and in turn drives the internal knowledge of companies. When employees change employer they will take the internal knowledge with them towards the new employer. The second erosion factor is the venture capital market. Since 1980 the venture capital market has increased significantly, paving the way for start-up's. In turn these start up's attract employees from established companies, getting a glimpse into their internal knowledge. However these start-up's also offer an interesting source of knowledge and innovation for these established companies. This brings us to the third erosion factor: external options for ideas sitting on the shelf. If the company cannot internally bring a new research result to market it can use start-ups by selling the IP, or even create a spin-off to bring the product to market. Lastly Chesbrough identifies the increasing capability of external suppliers as an erosion factor. With this he points towards the effect of all previous erosion factors, and more generally the innovation advancements, within other firms. These developments allow firms to create great capacity in their respective fields, thus operating as an external supplier at a level that an internal department could not possibly match. Chesbrough goes on to state that there erosion factors undermine the logic of Closed Innovation and break the virtuous circle of fundamental technology breakthroughs as talked about previously.

If Closed Innovation is no longer an option due to the erosion factors or the innovation scene within a certain industry firms should adopt a new approach, the Open Innovation model (Chesbrough, 2003). The Open Innovation model describes how firms can source ideas and ways of bringing these ideas to market both internally and externally. And with this Chesbrough (2003) states that firms can generate additional value, however these ideas should be supported by an according business model.

Insert figure 3 here

To make the distinction between Open and Closed Innovation clear Chesbrough (2003) highlights the key differences between the two. For example in contrast to Closed Innovation within Open Innovation firms must ensure a high influx of external ideas, encourage labour mobility, actively

engage in venture capitalism, pay attention to start-ups and recognize universities as a potential partner. Moreover he identifies some key mindset differences between the two models.

Empirical study

Data and general descriptive statistics

For this paper a database of publications in pharmaceutical journals was used. Information on this data was made available by the promoter of this master thesis, Linde Colen. The data was sourced via PubMed and was enriched with data from the CHI journal classification for data relating to the basicness of the publication. The classification of the basicness will be explained in more detail further on. All 59023 observations were published between 1994 to 2002, and are written by 62 pharmaceutical companies within Europe, the USA and Asia. These pharmaceutical companies were selected from the 2004 EU Industrial R&D Investment Scoreboard as being the largest R&D spenders (in absolute terms) in the pharmaceutical industry. The complete data gathering takes into account mergers and acquisitions of these 62 firms, by considering all subsidiaries owned by the firm in a particular year for at least 50% as producers of patents and publications for that firm.

For this study, I am mostly interested in two dimensions of a publication: the differences between independent research and research done in collaboration with others, and the difference between basic and applied research. .

The non-parametric variable 'publication type' gives us information on whether or not a publication is independent and can only take on two values: independent or collaborative. This variable has been determined by the affiliations mentioned on the publications under investigation. When the publication lists at least one affiliation of its others authors that is not an affiliation that can be linked to the firm or its subsidiaries, the publication is considered as being collaborative. Does the publication, on the other hand, only list affiliations of the firm or its subsidiaries, the publication can be considered as independent work of the firm.

The other dimension will be represented by the 'research type' variable. This variable indicates if the publication is a basic or applied research publication. Again this variable is non-parametric and can only take on of two values: basic research or applied research. In line with the literature the CHI-journal classification is used to determine this variable. as it seems that basic and applied research are published in designated journals, This classification is made on the basis of the journals where articles are published enabling a classification of journals' 'basicness' (Leten, Kelchtermans and Belderbos, 2010). However there is no real consensus of measuring basicness in such ways (Rosenbusch, Brinckmann and Bausch, 2011). Whilst an in depth analysis of this is not within the scope and relevancy of this particular research it is relevant to note that currently there is myriad of ways to innovation (e.g. Rosenbusch, Brinckmann and Bausch, 2011; Acs and Audretsch, 1988). For this study, the variable 'research type' considered an article as reporting on basic research when the CHI-journal classification has ranked the journal in which it is published in positions 3 or 4.

Methodology

As there is little research within the innovation literature on the current role of independent research within the pharmaceutical industry I will take on a non-directional hypothesis.

> **There is a difference between the characteristics of independent and collaborative publications across the pharmaceutical industry.**

In order to generate findings on the independent contribution of pharmaceutical firms to the publication stock I will first explore the possible differences by means of descriptive analysis. Most importantly, I will apply two group t-tests to examine the significance of mean differences between independent and collaborative publications. Two group t tests test the significance of the mean of two parent populations (Malhorta, Nunan and Birks, 2017; Serakan and Bougie, 2016).

The analysis will uphold a significance level of 0.05, unless stated otherwise. Some t test for example only reported a slight significance in which case the significance level is no higher than p=0.10. Because of the fact that our hypothesis is a non-directional one, I will mostly be testing the two tailed probability of our observations unless it's stated otherwise.

However the methodology does have some limitations and it should be noted that a t test preferably is conducted only when the number of observations is lower. The two populations or groups should also have same variance, however I have adjusted for the differences in variance within our t testing. In some cases the data is not normally distributed due to the fact that the variable under investigation only has few values that it can possible take, for example the level of basicness. In those cases however there is still an ordinal scale, which means that a t test remains a possible.

Insert table 1 here

Table 1 provides us with some initial descriptive statistics of the variables that are of interest for this research. The variables will be explained in their respective sections.

Independence of publications

When looking at the independent publications in the dataset contains 7399 publications. Some firms have a relatively higher stake in the dataset, but this is in line with what can observed in the entire dataset. Roche and Merck Co. both make up more than 10% of the independent publications in the dataset whilst Novartis, Eli Lilly, Bristol Myers Squib, GlaxoSmithKline and Pfizer all make up at least 5% of the independent publications each. There is also a difference in the year of independent publications as at least 80% of the publications are form the first 4 years of the dataset (1994-1997) whilst less then 20% of all independent publications are from the most recent years (1998-2002). This same discrepancy in years is not present in the entire dataset including collaborative publications.

Basicness level of the publications

Graph 4 describes the relative distribution of basicness level of the publications. This graph presents the basicness level of the publication as obtained from the CHI-journal classification from 1 to 4, moving from applied research to basic research. When comparing collaborative and independent publications we can already see that the independent publications are skewed slightly more to the right. This indicates already that the independent publications are slightly more basic compared to the collaborative publications.

Insert figure 4 here

We see the same pattern when we compare the mean level of basicness between independent and collaborative publications within the data. The mean level is higher among the independent publications, indicating again that the independent publications are more likely to be basic research publications compared to collaborative publications. When using a two sample student's t-test, shown in Table 2, we find that this difference is in fact a significant one. This finding is surprising. Because universities, the main generator of basic research, are a key collaboration partner in publishing, the expectation was that the collaborative publications would be comprised of more basic research. Furthermore it is contradictory to the concept and principals of absorptive capacity, the current understanding is that firms use their scientific capacity to access up-stream basic research (Cockburn and Henderson, 1998). The literature also argues that firms are conducting less basic research because of a narrower scope (Arora, Patacconi and Belenzon, 2012) also feeding the expectation of different findings.

However it is important to note that although the difference is significant it is at itself not that big.

One explanation for these findings can be found in the paper of Lim (2004), who argues that the industry context of the pharmaceutical research makes it particularly bound to both basic and applied research. So firms that conduct more basic research can absorb it better, this according to Lim is especially important in drug development because the innovations are depended on basic scientific discoveries. However this only can explain why both research types have a higher degree of basicness and still does not explain the heightened level of independent basic research.

One possible explanation for these findings could be that firms are actually conducting basic research on topics aligned to their own strategic core with the eye on further developing the research in to applied research and thus drug development. It is possible that for these very specific research topics firms fail to find research partners and in fact the difference is explained not by a choice of firms to conduct the research independently. In that case the difference is cause by a 'necessary evil' of firms in the sense that they must conduct the research but fail to interest any partners to co-develop the research. Furthermore by using internally developed basic research as the fundament for drug development they have more control over the intellectual property through the basic publication. Furthermore the firms might opt to develop the basic knowledge internally to speed up the total process of drug development.

Scientific quality of publications

Another topic of interest is the scientific quality of the independent publications. As mentioned before we look at citation of these publications to in this case determine the scientific value of the publications, as other scholars have done before (e.g. Cockburn and Henderson, 1998; Arora, Belenzon and Patacconi, 2015). However this is not without its limitations and it should be noted that it is used as approximation of scientific value. In the following part we will solely talk about citation to the publications by other scientific publications in journals, under the section further development we will discuss citation of the publications by patents and clinical projects.

Insert table 3 – 5 here

Surprisingly the amount of citations received is lower for independent publications, with collaborative publications receiving 25.959 citations on average and independent publications 22.886 on average (table 3). This is a significant difference ($p < 0.000$). When we sort the observations according to research type (table 4) we see that this difference is less outspoken among the basic research publications. Still collaborative publications have the upper hand but the difference shirks to 1.628 as apposed to the difference of 3.0731 among all publications. However this also has an impact on the significance as the difference on the mean amount of citations received between collaborative and independent publications among applied research publications is only slightly significant ($p = 0.508$). Naturally then when we look at the applied research publications the difference between the publication types is larger, with collaborative publications receiving 25.701 citations on average and independent publications 13.632 on average.

Sorting the observations according to publication type (table 5) we find no significant difference within the collaborative publications. However among the independent publications the difference in received citations is significant for basic and applied research($p = 0.000$). We see that basic independent publications receive more (24.418) citations on average as opposed to applied independent research publications (13.632).

So we find that across the board independent publications receive less citations, with independent applied research receiving the least citations of all. This can indicate that the independent publications have less scientific quality and that of the independent publications the basic research publications have the most scientific value. This could again indicate that independent internal research departments favour basic research and are in fact also more capable in basic research.

Further development

Patent development

To get a better understanding of independent contributions of pharmaceutical companies in the publication stock we look at how often the publications are cited other publications and how often they are used in clinical projects. Within the data there is information about the amount of references to the publication in patents and clinical projects by the company itself but also by other entities.

Insert table 6 - 8 here

Looking first whether the publication is mentioned as a reference on a patent of any of the 62 firms within the data (table 6) we see that there is a significant difference between collaborative and independent publications. The independent publications on average are cited more on patents.

However if I only looked at patents generated by the publishing firm itself (table 7) we see that this observation doesn't appear, on the contrary the average amount of firm patents generated by independent publications is slightly lower when compared to those from collaborative publications. This slight difference however is not a significant observation when the means are compared in a t test ($p = 0.642$).

Table 7 looks at the amount of firm patents relative to all patents, in essence looking at the likelihood of a patent being of the publishing firm. This gives a better indication of whether there is a difference in the patenting behaviour of firms when it comes to collaborative and independent publications. By making the amount of patents a relative variable I eliminated all observations without any patents and also decrease the impact of observations that might have unmatched amounts of patents related to them. Surprisingly this relative amount of firm patents is higher for collaborative publications, albeit at only a slights two-tailed significance ($p = 0.0942$). On average 5.39% of patents citing collaborative publication are of the publishing firm, whilst for independent publications this is 4.40% of all patents on average.

We observe thus that independent publications are more likely to lead to a patent, these patents however are less likely to be a patent by the publishing firm. When leading to a patent, it is less likely for an independent publications that the patent will be from the publishing firm

Insert table 9 - 10 here

To eliminate the chance of these previous findings being caused by missing data within the observation I eliminated all observations reporting no citations from patents. From the original 59024 publications this leaves us with 6825 observation. We immediately observe a higher mean amount of patents, on average a publication in the dataset gets referred to by 2.742 patents when we ignore all publications that do report zero patents. The mean amount of patents is higher for independent publications (2.864 patents referring to the publication) and lower for collaborative publications (2.717 patents referring to the publication). So we still observe a higher mean for independent publications as in table 6, contrary to the analysis with the entire dataset, here the difference is not significant when using a t tests.

Looking again only at the patents form the publishing firm itself we see that even when eliminating all observations reporting no citations from patents the collaborative publications still on average produce more patents of the firm. Independent publications only appear on 0.101 patents on average, collaborative publications on the other hand appear on 0.113 patents on average. As we already have an indication of the direction the difference would take we look at a one tailed significance test, and we can report that the difference here is only slightly significant ($p = 0.068$).

So for patent development we see that in absolute terms we can not find any significant differences, except a slightly significant difference in mean amount of patents of the firm citating the publications which is higher for collaborative publications, if we drop all observations reporting

zero patents referring to the publication. This slightly significant trend carries over when looking at the relative amount of firm patents citating the publication to all patents. There, a higher mean can also be observed at a slightly significant level. These findings could imply that firms do not a heightened commercial or technological interest in independent publication, as if this was the case we would expect the relative amount of firm patents to be higher for independent publications. Furthermore this indicates that at least when measured trough patents these publications are not any more disruptive then for example collaborative publications.

Project development

Insert table 11 - 13 here

Next to patents the clinical projects referring to the publications are also of interest. The first finding is the big difference in the mean amount of clinical projects referring to the publications when they are from collaborative or independent publications. The average amount of clinical projects of independent publications is almost double the average amount referring to collaborative publications (table 11). This difference is a significant observation, two tailed ($p < 0.000$), when tested with a two group t test. However just like with the patents this observation does not hold when we look at the clinical projects of the publishing referring to its own publication as seen in table 12. Here the difference is even reversed, with firms referring more to their own collaborative publications instead of independent publications. This slight difference however is not observed to be a significant one.

Following the previous analysis of patents the same procedure was made for clinical projects, creating a variable indicating the amount of firm clinical projects referring to the publication relative to the total projects. The same contradicting observation can be seen here. The relative average amount of clinical projects is higher for collaborative publications as opposed to independent publications. With this differences we can report a two tailed significant difference ($p = 0.014$), the significance is also higher with a one tailed positive difference ($p = 0.007$).

Insert table 14 - 15 here

To again eliminate the chance of these findings being caused by missing data within the observation I eliminated all observations reporting no citations from clinical projects. The difference in average amount of clinical projects referring to the publications when it is collaborative or independent becomes much smaller here. Independent publications still have a slightly smaller mean amount of clinical projects related to them (1.702 clinical projects on average), the difference with collaborative publication however is minimal (1.615 clinical projects on average for collaborative publications). This also shows when testing for significance, as we cannot report that this difference is a significant one.

I do find however a slightly significant difference ($p = 0.083$) between collaborative and independent publications when it comes to clinical projects of the firm when dropping all observations reporting zero clinical projects. In line with the analysis of the entire dataset we still observe a higher average amount of firm clinical projects from collaborative publications. When reporting only a one tailed significance this difference is even significant in itself ($p = 0.041$).

For project development again it is a nuanced story, similar to the analysis of patent development. Only here we observe to absolute amount of clinical projects referring to the publications to be in favour for independent publications. However this amount both in absolute and relative terms is in favour of the collaborative publications when looking at the firm clinical projects referring to the publications. This could indicate that the independent publications have a role in more clinical projects however they are not of higher importance to the publishing firms when it comes to clinical development.

Patent and project development

Insert figure 6 and 7 here

To further analyse the patent and project development following the publications in the dataset we again look in to the differences between collaborative and independent publication and the patents and clinical projects they bring forward. However now the observations are sorted by their research type, so depending on the fact if they publications in the dataset are basic research publications or applied research publications.

Looking at the graph of figure 6 we can see that on average the basic research publications get cited to the most, and within that research type the independent publications get cited slightly more on patents. This is contrast to the applied research publications, here the independent publication get cited slightly less.

On graph 7 we see the average amount of clinical project referring to the publication in the dataset by publication type and sorted by research type. Here the independent publications top the collaborative publications, in both research types. Overall the applied independent publication get referred to most in clinical projects, with 1.888 paten projects per publication on average. Interestingly the average amount of clinical projects for applied research publication is lower to the basic research publications when looking at just the collaborative publications.

It should be noted that these graphs do not include information on the missing data, and discards all observations that report zero patents or clinical projects.

Insert table 16 - 19 here

When testing these differences however the analysis does not yield a significant difference of amount of patents or even clinical projects when comparing the publication types. However the results of the mean amount of patents form the publications depending on publication type differs highly between the research types. This also shows in the data when we reverse the analysis and sort by research type (see table 17). The differences between basic and applied research are significant when looking at the patents in both independent ($p = 0.026$) and collaborative publications ($p < 0.000$). If we look back at table 16 we can see this quite clearly as there is an outspoken difference in the table of both research types. The amount of patents referring to the publication is highest for publications of basic research, and it is even higher when this publication is a basic research publication that was developed independently. This difference however is, as already reported, not significant.

When making the same comparisons with patens of the firm (table 18) we observe somewhat different results. The average amount of patents of the firm referring to the publications are higher with collaborative publication with both research types. When solely looking at applied research the difference between the publication types is significant (p = 0.044). Here the differences between the mean amount of firms patents of the publication types are however insignificant (p = 0.439 for the difference in collaborative publications and p=0.221) for the independent publication).

Insert table 20 - 22 here

We also applied this sorting of the observation according to research type to the clinical projects. When sorting on publication type we find no differences between collaborative and independent publication within both applied and basic research (table 20). The only significant difference this analysis yields really is when sorting per research type we find that collaborative publications are cited more by clinical projects when only considering applied research publications (p = 0.0186, see table 21). Furthermore there really are no significant differences regarding citations by clinical projects (see table 20 and 22).

These sorted analyses give us some additional insight in the patent development. Firstly we find that when sorted per publication type basic research publications are referred to more often on patents for both independent and collaborative publications, with a statically significant difference. Whilst when sorting on research type there is only a significant difference among the applied research publications, with collaborative publications being referred to more by patents. Furthermore we looked at clinical projects and the results of these analyses yielded only one significant difference. Overall the theoretical impact remains unchanged and fits with what is theorized at the ends of the paten development and project development sections.

Forward citations

In addition to information regarding the amount of patents and clinical projects referring to the publications in the dataset, it also contains the number of forward citations received by the publication. This includes information on citations of patents referring to the publications in the dataset. Furthermore the novelty in knowledge sources of patents referring to the publications and the novelty of the further development of patents referring to the publications. These variables can potentially yield a better understanding of the publication with regards to their own novelty and disruptiveness. Furthermore analysing the differences of these variables depending on independent and collaborative publications and basicness level might yield a better understanding of independent publications of pharmaceutical firms.

Insert table 23 and 24 here

When making the distinction between just collaborative and independent publications we only find slightly significant difference between the average number of citations of patents referring to the publications and the average novelty in knowledge sources of patents referring to the publications. The average number of citations of patents referring to the publications is higher for independent publications. However the average novelty in knowledge sources of patents referring to the publications are lower for independent publications.

Insert figure 8 here

Looking graphically at the average number of citations of patents referring to the publication we see some differences when sorting on the basis of research type. The means of basic research and applied research are further apart from each other within collaborative publications, but there is nearly no difference between the means within the independent publications. The collaborative applied research publication have a higher mean then both independent publication types, and the collaborative basic research publications have a lower mean then both independent publication types.

Insert table 25 and 26 here

If we concentrate for a moment on the average number of citations of patents referring to the publications and sort the observations according to the research type we find some interesting differences. Among the basic research publications we find an significant difference ($p = 0.027$) in the average number of citations of patents referring to the publications, in favour of the independent publications. We can see that on average references to the independent publications get referred to by 16.757 patents, which is 1.593 patents more than the collaborative publications.

If we however combine the numbers of the publication types we can see that there is a difference among the research types. When we reverse our last analysis and sort they t test by research type and test for differences of on the average number of citations of patents referring to the publications among applied research and basic research publications we see that only one reports a significant difference. This difference is within the collaborative publications ($p = 0.018$). Here the forward citations is higher for applied research publications, with on average 15.163 patents referring to publications citing the original publication for basic research and 17.469 for applied research publications.

Insert table 10 here

Furthermore there is a significant difference in het maximum number of citations of patents referring to the publication between independent and collaborative publications. The average maximum amount is higher for independent publications, and this difference is significant in a two-tailed two sample t test ($p = 0.010$).

Insert table 27 - 30 here

We cannot report any significant differences in variables regarding novelty of patents referring to the publication and novelty in further development of patents referring to the publication when comparing the means of theses variables whilst making the distinction between independent and collaborative publications.

Whilst we observe some differences regarding the forward citations of independent publications the findings are mainly limited to the difference in amount of citations of patents referring to the independent publications. We cannot draw any conclusions in regards to novelty of the independent publications, nor does it indicate any reasons for these publications to be independent or collaborative.

To recap we find indications that the patents referring to independent publications are being referred to more by other patents. However we also see that the average novelty in knowledge sources of patents referring to the publications is lower for independent publications. Lastly we find that the patents referring to independent publications have a higher mean maximum amount of citations. The fact that these patents are cited more whilst the average novelty in knowledge sources is lower could indicate that the original publications have an effect in the novelty and/or disruptions of the patents. This could indicate, as we will argue later on, that these independent publications fill very specific gaps within the literature.

Development speed

A third general topic of interest within the dataset is the time it takes for publication to be referred to by patents. Firstly we made a simple comparison of these variables with the distinction between independent and collaborative publications.

Insert figure 9 here

Looking at the average time between publications and referring patents we see that the collaborative and independent publications follow a similar looking distribution. However the collaborative publications are skewed more to the right, indicating a higher concentration of publications that are referred to within 3 to 4 years of the publication. It should be noted that this graph does not consider references from a patent to a publication that has been published after the application of the patent.

Insert table 31 here

Looking at the mean of the average time between publications and referring patents we observe a significant difference, whereby the time is longer for independent publications. On average the time between publication and referring patents is 2.538 years for collaborative publications and 3.431 years for independent publications. When testing this difference for significance we find it to be a significant observation ($p < 0.000$).

Insert table 32 and 33 here

If we only consider the applied research findings for a moment we see an increase in the average time difference between collaborative and independent publications. The difference here is 0.962 years, up almost ten percent when we consider the entire dataset where the difference is just 0.892 years. Again the two group t test show a significance with a probability lower than 0.000. The difference in average time between publication and patent by a reference however decreases a bit when we look at only the basic research publications compared to the entire dataset. Within the basic research publications the average time between publication and patent reference is still higher for independent publications then for collaborative publications at a significant level ($p < 0.000$).

If however we were to group the test for differences in the mean according to the publication type and look for differences between basic research publications and applied research publications we do not find a significant difference in the average time between publications and referring patents.

Furthermore we looked at the maximum, minimum and average time between publication and reference on a patent, both including and excluding references before the publication. These analyses however did not yield any extra insights so they will not be discussed here.

Regression analysis

To test the differences between independent publications and collaborative publications in more depth we also include some regression analyses here. This will allow us to include the effect of other variables in the context of previous analyses. We will use these regression analyses to look at the relationship between some dependent and independent variables that are of interest, whilst controlling for differences in for example research investments, year, and geographic location. It should be noted that with all regression analysis the degree of association between variables does not necessarily imply any causality.

Insert table 34 here

This table show the descriptive statistics for the variable used in the regression analyses, including mean, minimum, maximum, standard deviation and amount of observations.

Explaining the likelihood on working independently

Insert table 35 here

Firstly we look into independent variables that might explain contributing that a publication is an independent one. For this we make use of a probit regression, as the variable 'publication type' (indicating whether or not a publication is independent) is a binary one. The predictor variables of interest here are whether a publication is basic or not, if it received government support, the number of authors, the year of publication, the continent of the publishing company, the R&D investments in million dollars in the year of publication, the total amount of clinical projects of a firm relative to the R&D investments and lastly the number of publications of the firm relative to the R&D investments in million dollars both in the year of publication. Furthermore we have corrected for the effects of the publishing firm and the effect of the world of science journal category within life sciences, however these results are not included in the table as not to clutter the table with these corrections.

In line with previous findings in this paper the probit regression confirms the positive relationship between basicness and publication type. The regression shows a significant effect on these variables, when the publication is basic the chance of it being an independent publication increases. For government support, number of authors and publication year we find a negative relationship. Furthermore we find no significant effect when looking at the location of the publishing firm's headquarter. We do find a relationship between R&D investments and a publication being independent, however the observed effect is rather small. Furthermore the total number of publications of the firm relative to the R&D investments in million dollars has a positive significant effect on the publication type. Lastly the regression also find a slightly significant positive effect of

the number of clinical projects of a firm relative to the publishing firm investments in research in million dollars on the likelihood of a publication being independent.

Insert table 36 here

Secondly we look at relationship between the publication type (independent or collaborative) and the amount of clinical trials referring to the publications as the dependent variable. For this we use a negative binominal regression unlike with the previous regression our dependent variable is no longer binary. This regression includes all the same variables as the previous one, however it does not directly correct for the effects of the publishing firm and the effect of the world of science journal category within life sciences. As the effects of the firm has been corrected trough research investments and geographical locations and we are interested in the entire industry instead of on firm level. This last reasons is also why we did not correct for the WoS journal category within life sciences, as we want to observe the effect on the entire pharmaceutical industry without taking categories in to account.

In contrast to our previous analyses we find here that the when a publication is basic it increases the likelihood of a higher amount of clinical project referring to the publications. Furthermore we observe a similar increase in likelihood with publication that are basic, and with publications that have a higher amount of authors. Again we see a negative relationship between the dependent variable and the year of publication, so the more recent publications get referred to les by clinical projects. This last finding however can also be caused by a time lag in our data.

Insert table 37 here

When running a linear regression with the average citation of patents referring to the publications as dependent publication we find no significant effect of the publication being independent or not. This model follows the example of the previous model and includes the same variables. This strengthens our previous analysis and indicates that indeed the fact if a publication is independent does not influence the citation of patents referring to the publication. This of interest as if it was somehow the case this would have an effect of the theoretical part of our discussion. For example this could have indicated that in fact independent publications have an increased commercial or technical value relative to other publications, however this does not appear to be the case.

Insert table 38 here

For this research we have also looked at the effect of a combined variable sorting the observation in 4 categories on the basis of whether they are basic or applied publications and on if they are independent or collaborative publications. The model took applied collaborative publications as a reference but we again did not find any significant difference related to the publication type.

Discussion

Our research provides new insights in the independent scientific publication output of pharmaceutical firms. We find evidence to support our hypotheses that there is a difference between the characteristics of independent and collaborative publications across the pharmaceutical industry. Firstly we find that these publication are more often basic research publications in comparison to collaborative publications. Whilst the majority of collaborative publications are also basic publications the average level of the independent publication is significantly higher. Furthermore our evidence suggest that the likelihood of an publication being independent diminishes if the publications is more recent or when the number of authors increases and lastly when there is government support for the publication. Moreover we find differences in the quality, measured by citations received, of independent publication is trailing behind when compared to collaborative publications. We also find a difference regarding this quality within independent publications itself, as basic independent publication receive significantly more citations then independent applied publications.

Secondly we find differences in the referencing to independent publication by patens and clinical projects when compared to collaborative publications. The first finding in this regard is that relative to all citations by patents collaborative publications get cited more by their own publish firms, on a slightly significant basis. However we do not report any significant differences in the general average of patents and firm related patents referring to independent publications. Neither are we able to report any differences regarding clinical projects other than a higher amount of average clinical projects and average clinical projects by the publishing firm referring to collaborative publication. These differences however are just slightly significant. The relative mean amount of firm clinical projects however differs significantly, with collaborative publications being cited more on patents by the publishing firm relative to all patent citations when compared to independent publications.

Further we looked at the attributes of the patents referring to the independent publications and testes them for differences in comparison to the patents referring to collaborative publications. We reported higher average number of citations of patents referring to the publication for independent publications, at a slightly significant level. If we look at only the basic research publication we see that independent publications again have a higher average amount of citations, and we find this difference to be statistically significant. This same difference cannot be observed when solely looking at applied research publications. However these findings are contested by our regression (table 37) where we find no effect of the publication type on the number of citations of patents referring to the publications. Moreover we report a higher average novelty in knowledge sources of patents referring to the publications for collaborative publications, again at a slightly significant level. Finally we also observe a significant difference in the maximum number of citations of patents referring to the publication, where independent publication have a higher average maximum.

Lastly a significant difference is found in the average time between publication and reference to the publication on patents when comparing the two publications types. The average time is longer for

independent publication, this same trend continues when the observations are sorted according to research type.

We find thus that the independent research publications are more often basic publications. Our findings suggest that they are being referred to less by patents and clinical projects. Their forward citations are higher (on average and on maximum basis), however the novelty of within these citations is lower. And lastly the time between publication and reference on average is longer. As per the setup of this research, and it explorative nature, the questions we now ought to answer is what might be the possible reasons for these results be?

One possible explanation is that internal research departments, at least partially, publish more basic research because they need to fill in some very specific gaps in their own strategic core research field. It could be that because research and innovation is getting harder as science itself advances (Jones, 2009) and firms have an increasingly narrower scope (Arora, Belenzon and Patacconi, 2015; Jones, 2009) firms increasingly need to rely on their own independent research for basic, specific and early parts of the research process. This specificness could also explain why these publications are cited less by patents and clinical projects as other firms or institutes will not cite these publication because of a lack in the strategic or scientific fit. However we see that relative to all patent and clinical projects the average amount of firm patents and clinical projects is not higher for independent publication, which would be expected if this theory holds true. This expected effect however can be found in our second regression analysis (table 36). We find there that when a publication is basic it increases the likelihood of a higher amount of clinical project referring to the publications, we also find the same positive effect when it comes to publications that are independent. The longer average time between publication and reference by a patent could indicate that other institutions are not up to speed on the knowledge contributed by these publications.

Another potential explanation for these findings could be the influence of (star) scientist within the internal research department and their own inclination for basic research and certain research domains. Previous scholars have already explored the influence of researchers on the publication stock of private firms, as allowing them to publish is an argument to convince these research to work within private (Henderson and Cockburn, 1998). These research have very specific fields of expertise (Jaffe, 1986) and our results could be a results of these experts influencing the search of pharmaceutical companies. The direction of the search being in alignment with the inclination of the scientists instead of the strategic core of the company.

A third possible reasoning could be that our results are attributable absorptive capacity. We know that one reasons to conduct basic research is the positive effect it has on the absorptive capacity of firms (Cohen and Levinthal, 1989) and that there is a myriad way of ways that firms can exploit external research (e.g. Arora, Belenzon and Patacconi, 2015: Arora and Gambardella, 1990; Higgins and Rodriguez, 2005; Mowery, 2009). With the emergence of Open Innovation firms are increasingly relying on absorptive capacity for their innovation productivity (Chesbrough, 2003) and it could be that our findings are the results of the efforts of firms to increase and maintain their absorptive capacity. This reasoning is also supported by the findings in our first regression (table 35) where we find a positive relationship between a publication being independent and the R&D

investments of a firm and the total amount of publications and clinical projects of the firm relative to the R&D investments.

Lastly we would like to argument that the findings are caused by the search of pharmaceutical firms towards disruptive and breakthrough innovations. In line with the reasoning of our first explanation it could be that firms are research fundamental topics in order to further develop into disruptive innovations. This could also explain the positive effect of a publication begin independent on the average amount of clinical projects found in the regression analysis.

We conclude that in the current pharmaceutical landscape firms still independently contribute to the publication stock, we proposed some possible theoretical mechanism to explain the observed characteristics of independent research publications. These mechanism however could be co-existence but this should be the subject of further research. To get more insights in to the independent contributions of internal research departments to the publication stock within drug discovery we suggest future research could direct attention to the internal reasoning of these research departments and their patent and clinical projects.

The findings of these dissertation are limited firstly by the research aims, as we have only looked at the publications and their citations data. Furthermore it should be noted that this research is limited to the drug development industry and the sample data consists only of publications in Asian, European and North-American pharmaceutical companies from 1994 till 2002. Lastly although the sample itself was quite large data on some subclasses was limited in the analysis of this dissertation, and might not be representative for the industry at large.

Conclusion

This dissertation was set out to answer the research question: " For which research do internal research departments independently contribute to the publication stock within drug discovery?". This is pertinent question now that industries are shifting more to the open innovation paradigm, and at the same time investments towards research by private firms is on the decline.

Whilst the pharmaceutical sector has some very specific caveats when it comes to science and innovation there is still a link between the two. The literature study describes the well documented flow of R&D in the pharmaceutical sector and the discrepancies between basic and applied research. Furthermore scientific research is paired with some disadvantages. These being an uncertainty and a risk of failure, a time lag between investments and pay off and the chance of spill overs and free riders within the industry. Nevertheless firms still invest in research, because of its benefits. Firstly the absorptive capacity it creates, but also the productivity and profitability it can generate play a big role. Some scholars argue that in house scientific research also attracts more star scientist towards the firm. However in recent years there has been a decline of research investments within the pharmaceutical industry.

In this research I find support for the hypothesis that there is in fact a difference between the characteristics of independent and collaborative publications across the pharmaceutical industry. Our results indicate that independent research is more basic. Furthermore we see less independent research over time and publications with a higher number of researchers is more likely to be independent. Independent, and especially applied independent, publications receive significantly less citations of other research publications. Moreover we find differences in references to independent publications by patens and clinical projects. I also looked at the further development of the patents referring to the publications and we here found some differences as well. Lastly the results indicate a difference in the development times of patent referring to the publications between independent and collaborative publications.

I attribute these findings to three separate, although not mutually exclusive, reasonings. I firstly argue that in part internal research departments publish and thus conduct more basic research in order to fill some very specific gaps in the literature to complement their own core strategic research. Secondly we argue that the differences are cause by the influence of the internal scientists and their influence on the direction of internal research. The reasoning attributes the results to the effects of absorptive capacity. And the last reasoning states that the findings are cause by the search of firms towards disruptive and breakthrough innovations. These mechanism in our view could a solid direction for further research, as these could be of relevance to both public policy makers and the pharmaceutical industry itself.

References

Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: an empirical analysis. *The American economic review*, 678-690.

Adams, J. D. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of political economy*, *98*(4), 673-702.

Arora, A., Belenzon, S., & Patacconi, A. (2015). *Killing the golden goose? The decline of science in corporate R&D* (No. w20902). National Bureau of Economic Research.

Arora, A., & Gambardella, A. (1990). Complementarity and external linkages: the strategies of the large firms in biotechnology. *The journal of industrial economics*, 361-379.

Arrow, K. J. (1972). Economic welfare and the allocation of resources for invention. In *Readings in industrial economics* (pp. 219-236). Palgrave, London.

Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology. Harvard Business Press.

Chesbrough, H. W. & Bogers, M. (2014). Explicating open innovation: Clarifying an emerging paradigm for understanding innovation. *New Frontiers in Open Innovation. Oxford: Oxford University Press, Forthcoming*, 3-28.

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: the two faces of R & D. *The economic journal*, *99*(397), 569-596.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.

Cockburn, I. M., & Henderson, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The journal of industrial Economics*, *46*(2), 157-182.

Gambardella, A. (1992). Competitive advantages from in-house scientific research: The US pharmaceutical industry in the 1980s. *Research policy*, *21*(5), 391-407.

Gordon, R. J. (2012). Is US economic growth over? Faltering innovation confronts the six headwinds (No. w18315). National Bureau of Economic Research.

Griliches, Z. (2007). *R&D and productivity: The econometric evidence*. University of Chicago Press.

Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research policy*, *31*(4), 477-492.

Hicks, D., Ishizuka, T., Keen, P., & Sweet, S. (1994). Japanese corporations, scientific research and globalization. *Research Policy*, *23*(4), 375-384.

Higgins, M. J., & Rodriguez, D. (2006). The outsourcing of R&D through acquisitions in the pharmaceutical industry. *Journal of financial economics*, *80*(2), 351-383.

Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, 957-970.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. *NATIONAL BUREAU OF ECONOMIC RESEARCH,* Working Paper No. 1815.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?. *The Review of Economic Studies*, *76*(1), 283-317.

Kaitin, K. I. (2010). Deconstructing the drug development process: the new face of innovation. *Clinical Pharmacology & Therapeutics*, *87*(3), 356-361.

Lee, J. (2003). Innovation and strategic divergence: An empirical study of the US pharmaceutical industry from 1920 to 1960. *Management Science*, *49*(2), 143-159.

Leten, B., Kelchtermans, S., & Belderbos, R. (2010). Internal basic research, external basic research and the technological performance of pharmaceutical firms. *Katholieke Universiteit Department of Managerial Economics, Strategy and Innovations Working Paper*, (1003).

Lim, K. (2004). The relationship between research and innovation in the semiconductor and pharmaceutical industries (1981–1997). *Research policy*, *33*(2), 287-321.

Mansfield, E. (1980). Basic research and productivity increase in manufacturing. *The American Economic Review*, *70*(5), 863-873.

Mowery, D. C. (2009). Plus ca change: Industrial R&D in the "third industrial revolution". *Industrial and corporate change*, *18*(1), 1-50.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy*, *67*(3), 297-306.

Pavitt, K. (1991). What makes basic research economically useful?. *Research policy*, *20*(2), 109-119.

Reagan, M. D. (1967). Basic and applied research: A meaningful distinction?. *Science*, *155*(3768), 1383-1386.

Rosenberg, N. (1990). Why do firms do basic research (with their own money)?. In *Research Policy, 19,* (pp. 165-174).

Rosenbusch, N., Brinckmann, J., & Bausch, A. (2011). Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of business Venturing*, *26*(4), 441-457.

Salter, A. J., & Martin, B. R. (2001). The economic benefits of publicly funded basic research: a critical review. *Research policy*, *30*(3), 509-532.

Ward, M. R., & Dranove, D. (1995). The vertical chain of research and development in the pharmaceutical industry. *Economic inquiry*, *33*(1), 70-87.

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of management review*, *27*(2), 185-203.

Zucker, L. G., Darby, M. R., & Armstrong, J. (1998). Geographically localized knowledge: spillovers or markets?. *Economic Inquiry*, *36*(1), 65-86.

Appendix

Figures

Figure 1: A fully integrated pharmaceutical network model of drug development (Kaitin, 2010)

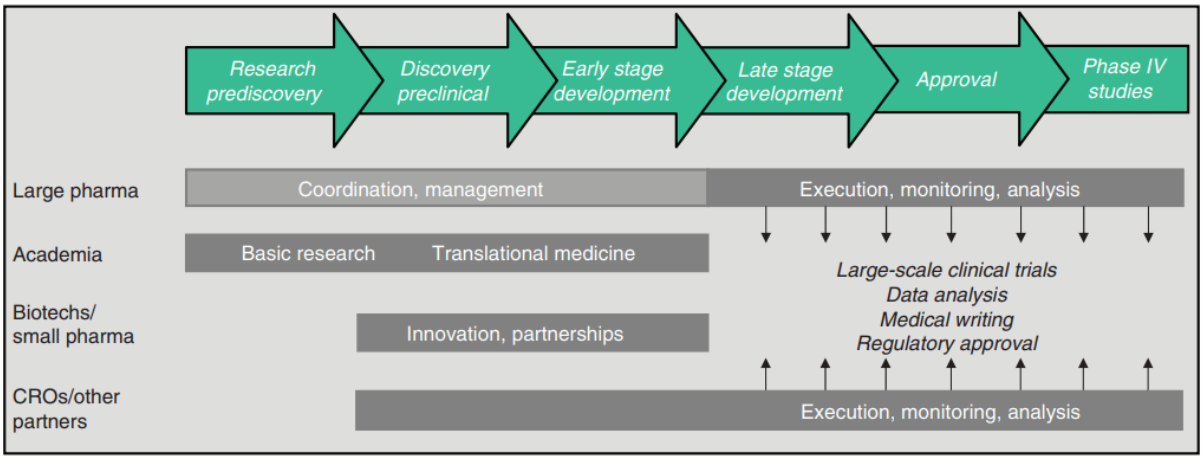
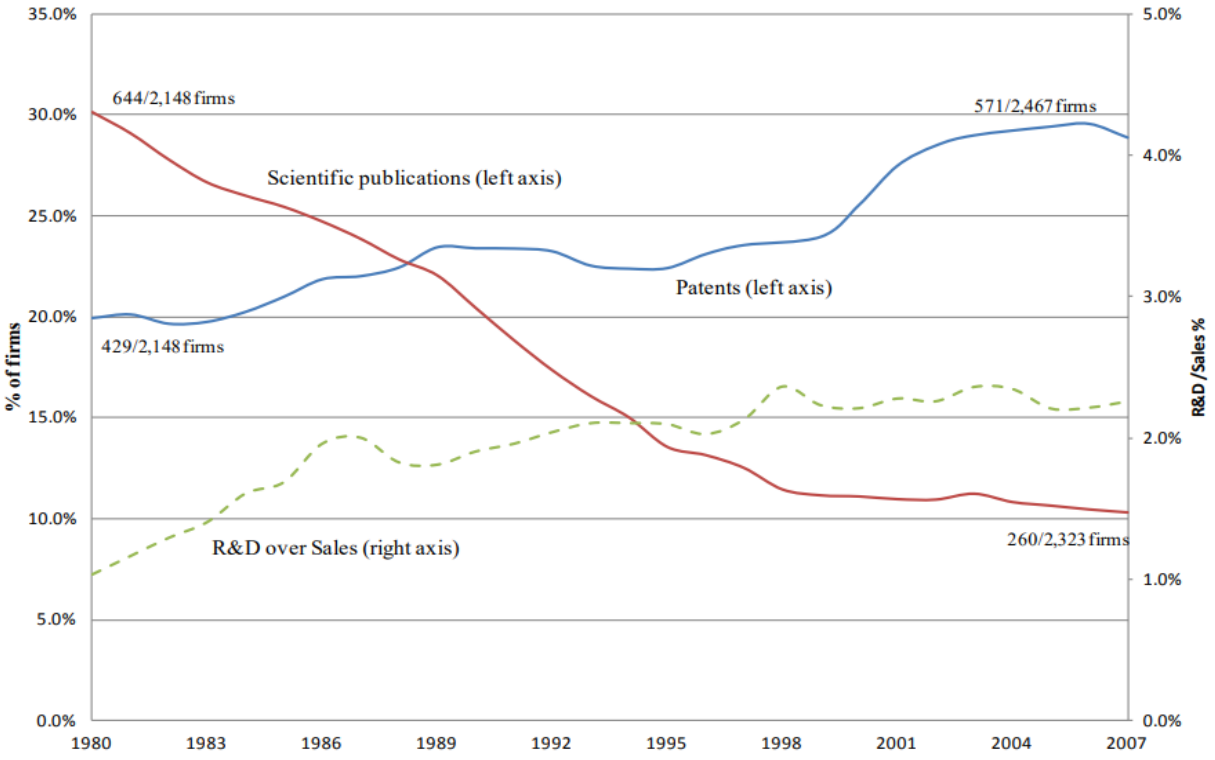


Figure 2: Investments in Science and Technology over time 1980-2007 (Arora Belenzon and Patacconi, 2015)



Note: This figure presents the share of publishing and patenting firms of all Compustat firms with at least one year with non-zero R&D expenditures, over time. Data source: Compustat, Web of Science, PatStat.

Figure 3: The open innovation model and closed innovation model (Chesbrough, 2003)

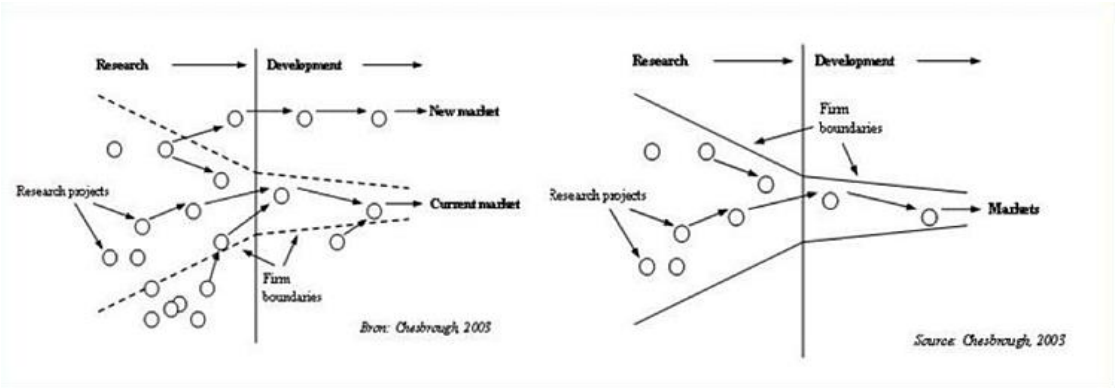


Figure 4: Relative frequency of publication level

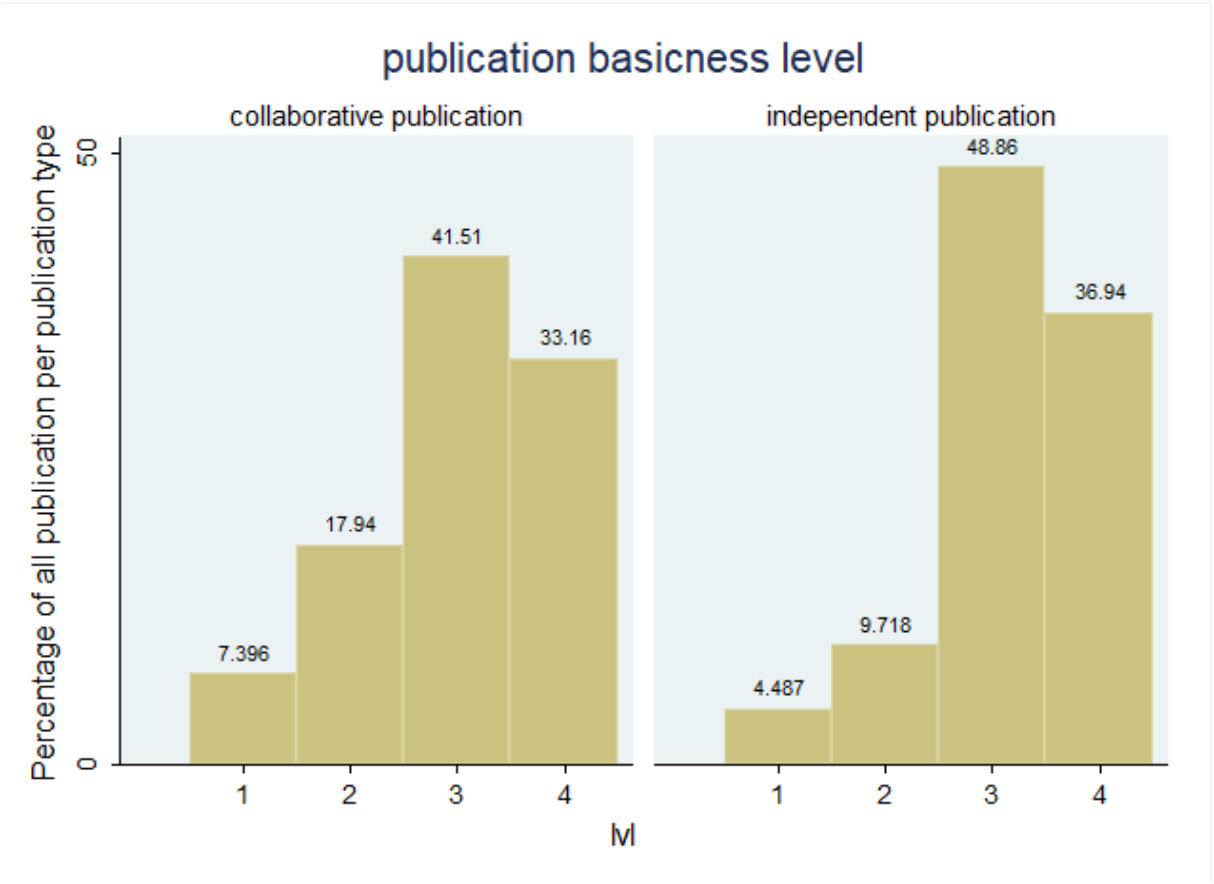


Figure 5: average basicness of each publication type

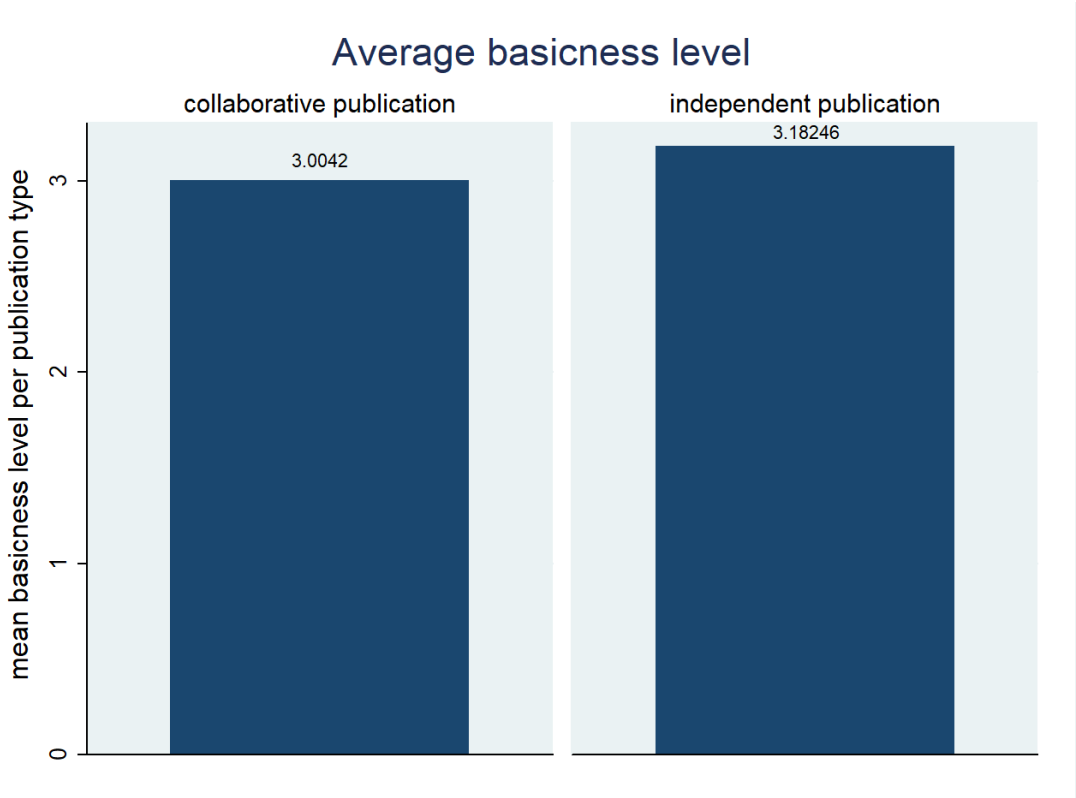


Figure 6: average amount of patents of each publication type sorted by research type

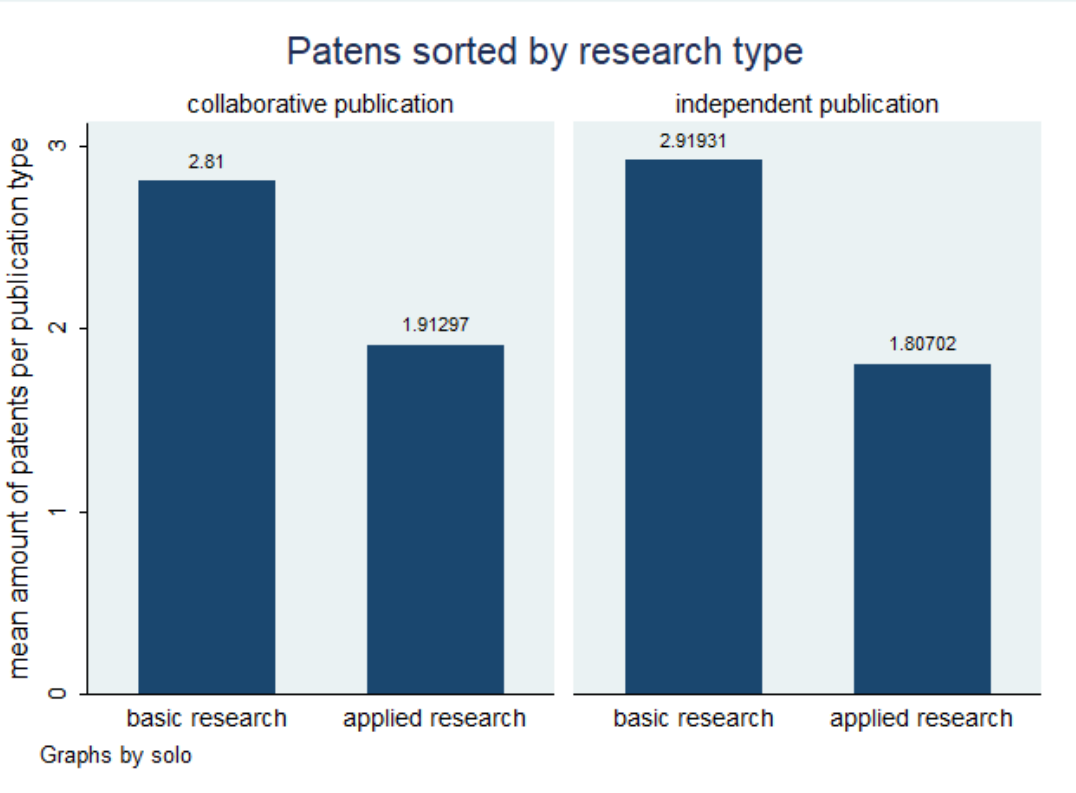


Figure 7: average amount of clinical projects of each publication type sorted by research type

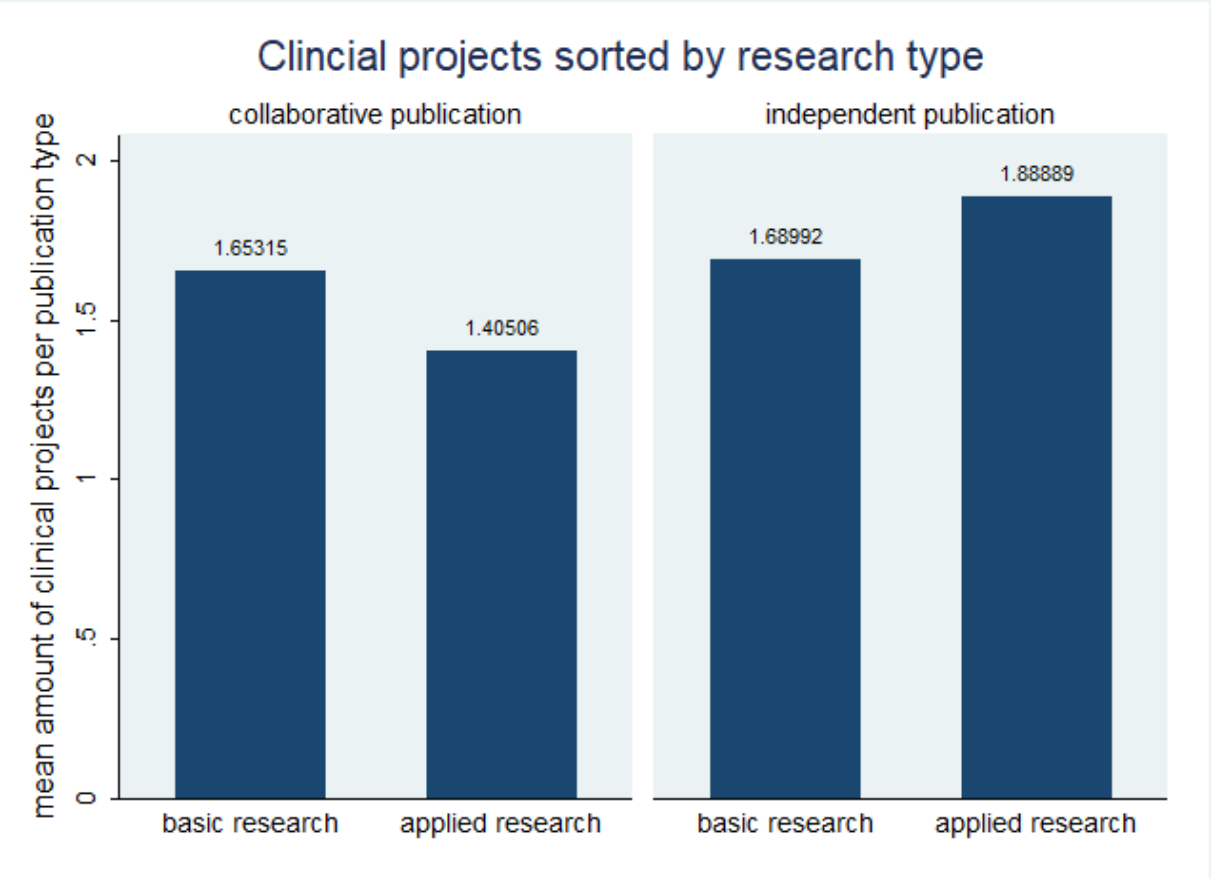


Figure 8: average number of citations of patents referring to the publication by publication type sorted by research type

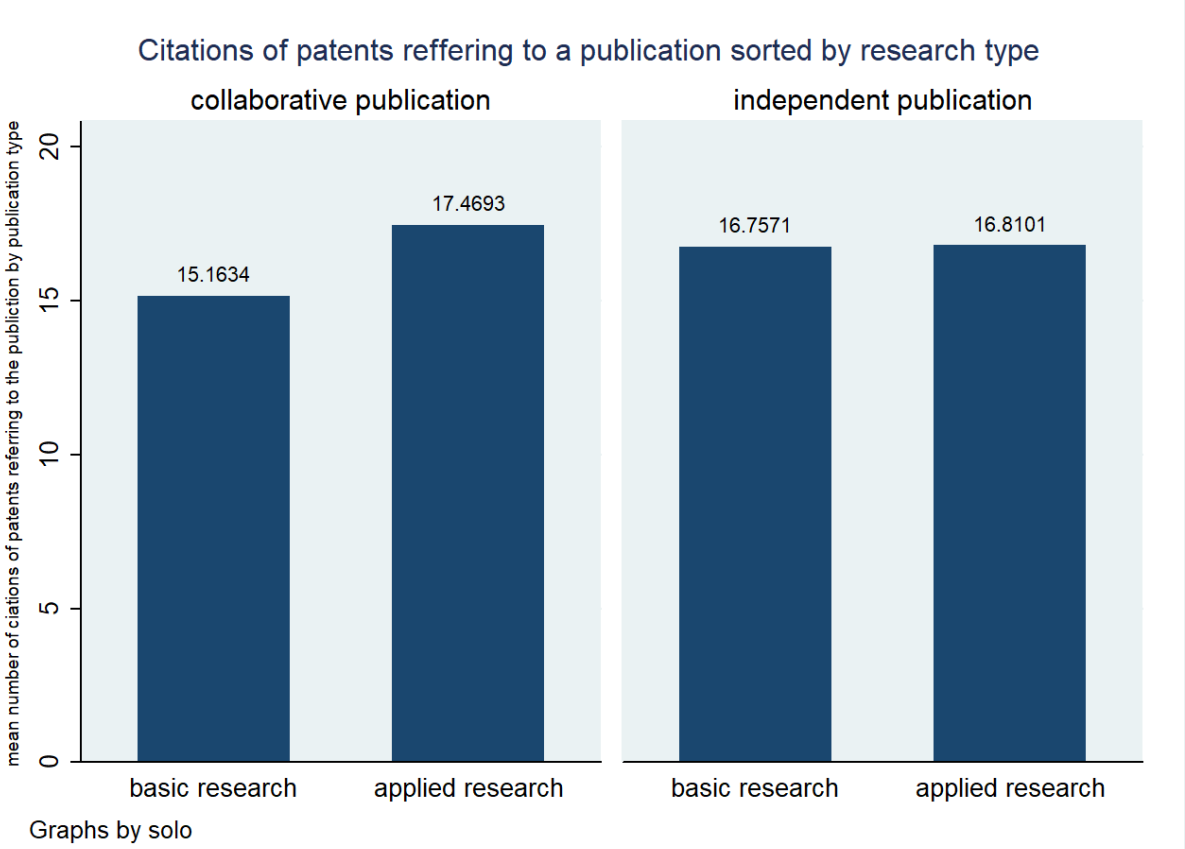
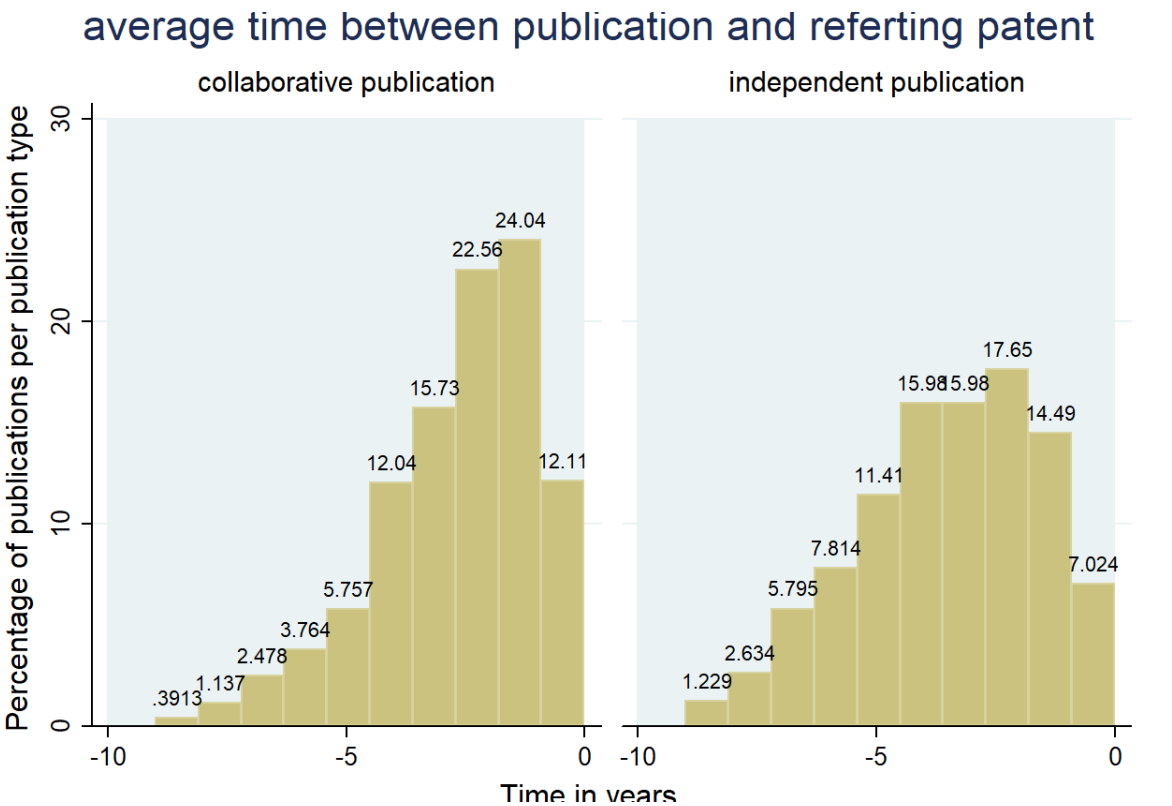


Figure 9: Average time between publication and referring patent after publication



Tables

Table 1: general descriptive statistics

Variable	Observations	Means	Standard deviation	Minimum	Maximum
Basicness level of the publication	59023	3.026549	.8865483	1	4
Research type (1: applied research, 0: basic research)	59024	.239428	.426738	0	1
Publication type (1: Collaborative publication, 0:Independent publication)	59024	.1253558	.3311247	0	1
Citations of the publication	59024	25.57429	73.31891	0	5370
Patents referring to the publication	59024	.3170914	1.925071	0	196
Patents of the publishing firm referring to the publication	59024	.0148245	.413125	0	94
Clinical projects referring to the publication	59024	.0182976	.2366056	0	9
Clinical projects of the publishing firm referring to the publication	59024	.0004913	.0236403	0	2
Citations of the patents referring to the publication	6825	15.63272	20.40241	0	281
Maximum citations of the patents referring to the publication	6825	24.00029	34.51391	0	391
Average novelty of the patents referring to the publication	6825	.2141228	1.072262	0	19
Maximum novelty of the patents referring to the publication	6825	.5991209	4.039576	0	112

Average novelty of sources used in patents referring to the publications	6825	.4.2521921	175335	0	445.5
Maximum novelty of sources used in patents referring to the publications	6825	8.953846	35.90901	0	889
Average novelty in further development of patents referring to the publication	6825	1.19839	8.147338	0	273
Maximum novelty in further development of patents referring to the publication	6825	2.50871	11.49292	0	273
Average time between publication and reference by a patent of the publication	6506	-2.694868	1.887227	-9	0
Maximum time between publication and reference by a patent of the publication	6506	-2.154012	1.897597	-9	0

Table 2: basicness level by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	51624	3.004203	.0039533	.8982275	2.996455	3.011952
Independent publications	7399	3.182457	.0091008	.7828317	3.164617	3.200297
Total	59023	3.026549	.0036492	.8865483	3.019397	3.00701
Difference		-.1782536	.0099224		-.1977034	-.1588037
t = -17.9648						
Satterthwaite's degrees of freedom =10400.3						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0000		Pr(|T| > |t|) =0.0000		Pr(T > t) =1.000		

Table 3: Average citation of scientific publications to the publications by publications type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	51625	25.95954	.3316724	75.35975	25.30945	26.60962
Independent publications	7399	22.88634	.662837	57.01554	21.58699	24.18569
Total	59027	25.57429	.30178778	73.31891	24.98279	26.1658
Difference		3.073199	.7411879		1.620344	4.526054
						t = 4.1463
						Satterthwaite's degrees of freedom = 11463.5
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) =1.0000		Pr(|T| > |t|) = 0.0000			Pr(T > t) = 0.0000	

Table 4: Average citation of scientific publications to the publications by publications type sorted by research type

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	38544	26.04717	.373733	73.3736	25.31464	26.77969
Independent publications	6348	24.4184	.7453427	59.38469	22.95728	25.87952
Total	44892	25.8184	.3377583	71.56333	25.15484	26.47886
Difference		1.628767	.8337938		-.0056395	3.263174
						t = 1.9534
						Satterthwaite's degrees of freedom =9837.43
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.9746		Pr(|T| > |t|) = 0.0508			Pr(T > t) = 0.0254	

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	13081	25.70132	.7076156	80.93155	24.31429	27.08835
Independent publications	1051	13.63273	1.189476	38.56177	11.29871	15.96675
Total	14132	24.80378	.6614663	78.63384	11.29871	15.99675
Difference		12.06859	1.384042		9.354195	14.78299
						t = 8.7198
						Satterthwaite's degrees of freedom = 1905.55

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 1.000	Pr(|T| > |t|) = 0.0000	Pr(T > t) = 0.0000

Table 5: Average citation of scientific publications to the publications by research sorted by publication type

Collaborative publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic research	38544	26.04717	.373733	73.3736	25.31464	26.77969
Applied research	13081	25.70132	.7076156	80.93155	24.31429	27.08835
Total	51625	25.95954	.3316724	75.35975	25.30945	26.60962
Difference		.3458443	.8002475		-1.222703	1.914392
t = 0.4322						
Satterthwaite's degrees of freedom = 20844.7						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.6672		Pr(|T| > |t|) = 0.6656		Pr(T > t) = 0.3328		

Independent publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic research	6348	24.4184	.7453427	59.38469	22.95728	25.87952
Applied research	1051	13.63273	1.189476	38.56177	11.29871	15.96675
Total	7399	22.88634	.662837	57.01554	21.58699	24.18569
Difference		10.78567	1.403705		8.032779	13.53856
t =7.6837						
Satterthwaite's degrees of freedom = 1985.79						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 1.0000		Pr(|T| > |t|) = 0.0000		Pr(T > t) =0.0000		

Table 6: average amount of patents from the publications by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	51625	.2981695	.0085483	1.942271	.281447	.3149242
Independent publications	7399	.4491147	.0208688	1.795077	.408206	.4900235

Total	59024	.3170914	.0079238	1.925071	.3015607	.332622
Difference		-.1509453	.0225517		-.1951511	-.1067394
t = -6.6933						
Satterthwaite's degrees of freedom = 10048.4						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) =0.0000		Pr(|T| > |t|) = 0.0000		Pr(T > t) = 1.0000		

Table 7: average amount of firm patents from the publications by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	51625	.0146634	.001927	.4378263	.0108866	.0184403
Independent publications	7399	.01559481	.001987	.1709182	.012053	.0198432
Total	59024	.0148245	.0017037	.4139125	.0114852	.0181637
Difference		-.0012847	.0027679		-.00671	-0041406
t = -0.4641						
Satterthwaite's degrees of freedom = 24722.7						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.3213		Pr(|T| > |t|) = 0.6426		Pr(T > t) = 0.6787		

Table 8: relative amount of firm patents to all patents by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	.0539907	.0026589	.2001221	.487783	.059203
Independent publications	1160	.0440064	.0053362	.1817451	.0335367	.0544762
Total	6825	.0522937	.0023863	.19714525	.0476158	.0569716
Difference		.0099842	.0059619		-.0017089	.0216774
t = 1.6747						
Satterthwaite's degrees of freedom = 1783.43						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9529		Pr(|T| > |t|) = 0.0942		Pr(T > t) = 0.0471		

Table 9: average amount of patents from the publications by publication type (missing data dropped)

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	2.717211	.0700638	5.273437	2.5798589	2.854563
Independent publications	1160	2.864655	.1084474	3.693586	2.65188	3.07743
Total	6825	2.742271	.0610075	5.040046	2.622677	2.861865
Difference		-.147442	.1291115		-.4006345	.105746
t = -1.1420						
Satterthwaite's degrees of freedom = 2248.3						
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.1267		Pr(|T| > |t|) = 0.2536			Pr(T > t) = 0.8732	

Table 10: average amount of firm patents from the publications by publication type (missing data dropped)

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	.1336275	.0174816	1.315773	.0993569	.1678981
Independent publications	1160	.1017241	.0123783	.421588	.0774379	.1260104
Total	6825	.1282051	.0146625	1.211319	.0994621	.1569482
Difference		.0319034	.0214202		-.0100884	.0738952
t = 1.4894						
Satterthwaite's degrees of freedom = 5729.24						
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.9318		Pr(|T| > |t|) = 0.1364			Pr(T > t) = 0.0682	

Table 11: average amount of clinical projects from the publications by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	51625	.016368	.0009803	.2227375	.0144466	.0182895
Independent publications	7399	.0182976	.0009739	.3165986	.024546	.0389761
Total	59024	.0182976	.0009739	.2366056	.0163888	.0202065
Difference		-.015393	.0038089		-.0228595	-.0079265
t = -4.0413						

		Satterthwaite's degrees of freedom = 8478.72
Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.0000	Pr(|T| > |t|) = 0.0001	Pr(T > t) = 1.0000

Table 12: average amount of firm clinical projects from the publications by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	51625	.0005036	.0001061	.0241013	.0002957	.0007115
Independent publications	7399	.0004055	.0002341	.0201333	-.0000534	.0008643
Total	59024	.0004913	.000973	.02364403	.003006	.000682
Difference		.0000982	.00257		-.0004055	.00060139
						t = 0.3820
						Satterthwaite's degrees of freedom = 10684.3
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.6488		Pr(|T| > |t|) = 0.7024		Pr(T > t) = 0.3512		

Table 13: relative amount of firm clinical projects to all clinical projects by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	523	.043021	.0087248	.1995302	.0258809	.0601612
Independent publications	138	.013285	.0084086	.0987792	-.0033425	.0299125
Total	661	.036129	.007136	.1834655	.0228009	.0508249
Difference		.029736	.0121173		.005923	.0535491
						t = 2.4540
						Satterthwaite's degrees of freedom = 452.989
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9927		Pr(|T| > |t|) = 0.0145		Pr(T > t) = 0.0073		

Table 14: average amount of clinical projects from the publications by publication type (missing data dropped)

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	523	1.615679	0.665675	1.522344	1.48906	1.746452
Independent publications	138	1.702899	.1358314	1.595658	1.434301	1.971496
Total	661	1.633888	.10597869	1.537117	1.516493	1.751284
Difference		-.0872198	.151266		-.3854344	.2109948
t = -0.5766						
Satterthwaite's degrees of freedom = 207.568						
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.2824		Pr(|T| > |t|) = 0.5648			Pr(T > t) = 0.7176	

Table 15: average amount of firm clinical projects from the publications by publication type (missing data dropped)

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	523	.0497132	.0102544	.2345104	.0295682	0.698582
Independent publications	138	0.217391	.0124591	.1463618	-.002898	.0463762
Total	661	.0438729	.0124591	.219538	.021277	.0606182
Difference		.0279741	.0161364		-.0037644	.0597125
t = 1.7336						
Satterthwaite's degrees of freedom = 344.04						
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.9581		Pr(|T| > |t|) = 0.0839			Pr(T > t) = 0.0419	

Table 16:average amount of patents from the publications by publication type sorted by research type (missing data dropped)

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5079	2.81002	.077167	5.499466	2.658721	2.961282
Independent publications	1103	2.919311	.112858	3.748178	2.69787	3.140752
Total	6182	2.829505	.0665185	5.230062	2.699106	2.959904
Difference		-.109309	.1737464		-.4499123	.2312943
						t = -0.6291
						Satterthwaite's degrees of freedom = 6180
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.2646		Pr(|T| > |t|) = 0.5293			Pr(T > t) = 0.7354	

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	586	1.912969	.1011668	2.448988	1.714275	2.111664
Independent publications	57	1.807018	.2869756	2.166618	1.232136	2.381899
Total	643	1.903577	.0955904	2.423927	1.715869	2.091285
Difference		.1059517	.3365455		-.5549131	.7668166
						t = 0.3148
						Satterthwaite's degrees of freedom =641
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.6235		Pr(|T| > |t|) = 0.7530			Pr(T > t) = 0.3765	

Table 17: average amount of patents from the publications by research type sorted by publication type (missing data dropped)

Collaborative research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic research	5079	2.81002	.077167	5.499466	2.658721	2.961282
Applied research	586	1.912969	.1011668	2.448988	1.714275	2.111664
Total	5665	2.717211	.0700638	5.27437	2.579859	2.854563
Difference		.8970327	.2297791		.4465776	1.347488
						t =3.9039
						Satterthwaite's degrees of freedom = 5663
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) =1.0000		Pr(|T| > |t|) = 0.0001		Pr(T > t) = 0.0000		

Independent research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic research	1103	2.919311	.112858	3.748178	2.69787	3.140752
Applied research	57	1.807018	.2869756	2.166618	1.232136	2.381899
Total	1160	2.8646555	.1084474	3.693586	2.65188	3.07743
Difference		1.1129993	.5008604		.129598	2.094989
						t = 2.2208
						Satterthwaite's degrees of freedom = 1185
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9867		Pr(|T| > |t|) = 0.0266		Pr(T > t) = 0.0133		

Table 18:average amount of firm patents from the publications by publication type sorted by research type (missing data dropped)

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5079	.1382162	.0194202	1.38402	.1001442	.1762881
Independent publications	11303	.1051677	.0129486	.430043	.079761	.1305745
Total	6182	.1323196	.016122	1.267605	.1007149	.1639244
Difference		.0330485	.0233412		-.0127094	.0788063
t = 1.4159						
Satterthwaite's degrees of freedom = 5545.86						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9216		Pr(|T| > |t|) = 0.1569		Pr(T > t) = 0.0784		

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	586	0.938567	.0150685	.3647687	.0642618	.1234515
Independent publications	57	0.350877	.0245882	.1856372	-.0141684	.0843439
Total	643	.088647	.0139164	.3528856	.0613197	.1159742
Difference		.0587689	.0288382		.0015853	.1159526
t = 2.0379						
Satterthwaite's degrees of freedom = 104.55						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9780		Pr(|T| > |t|) = 0.0441		Pr(T > t) = 0.0220		

Table 19: average amount of firm patents from the publications by research type sorted by publication type (missing data dropped)

Collaborative research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic research	5079	.1382162	.0194202	1.38402	.1001442	.1762881
Applied research	586	0.938567	.0150685	.3647687	.0642618	.1234515
Total	5665	.1336275	.0174816	1.315773	.0993569	.1678981
Difference		.0443595	.0574063		-.0681785	.1568976
t = 0.7727						
Satterthwaite's degrees of freedom = 5663						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.7801		Pr(|T| > |t|) = 0.4397		Pr(T > t) = 0.2199		

Independent research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic research	11303	.1051677	.0129486	.430043	.079761	.1305745
Applied research	57	0.350877	.0245882	.1856372	-.0141684	.0843439
Total	1160	.1017241	.0123783	.421588	.0774379	.1260104
Difference		.07008	.0572531		-.0422513	.1824114
t = 1.2240						
Satterthwaite's degrees of freedom = 1158						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.8894		Pr(|T| > |t|) = 0.2212		Pr(T > t) = 0.1106		

Table 20:average amount of clinical projects from the publications by research type sorted by publication type (missing data dropped)

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	444	1.653153	.0728052	1.534101	1.510067	1.79625
Independent publications	129	1.689922	.1360425	.1545146	1.420739	1.959106
Total	573	1.661431	.0641388	1.535317	1.535455	1.787407
Difference		-0.367693	.1542989		-.340969	.2674303
						t = -0.2383
						Satterthwaite's degrees of freedom = 206.914
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.40589		Pr(|T| > |t|) = 0.8119			Pr(T > t) = 0.5941	

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	79	1.405063	.162661	1.445763	1.08123	1.728897
Independent publications	9	1.888889	.7718024	2.315407	.1091093	3.668669
Total	88	1.454545	.1647558	1.545547	1.127075	1.782016
Difference	-.4838256	.788757			-2.276737	1.309086
						t = -0.6134
						Satterthwaite's degrees of freedom = 8.7247
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.2776		Pr(|T| > |t|) = 0.553			Pr(T > t) = 0.7224	

Table 21:average amount of firm clinical projects from the publications by research type sorted by publication type (missing data dropped)

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	444	.0427928	.10101308	.2134683	.0228825	.0627031
Independent publications	129	.0232558	.0133214	.1513025	-.0031029	.0496146
Total	573	.0383944	.0084052	.2011997	.0218855	0.549033
Difference		.019537	.016736		-.013402	.052476
						t = 1.1671
						Satterthwaite's degrees of freedom = 290.767
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.8780		Pr(|T| > |t|) = 0.2440			Pr(T > t) = 0.1220	

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	79	.0886076	.0368769	.3277692	.0151913	.1620239
Independent publications	9	0	0	0	0	0
Total	88	0.0795455	0.0332087	.3115254	.0135395	.1455514
Difference		.0886076	.0368769		.0151913	.1620239
						t = 2.4028
						Satterthwaite's degrees of freedom = 78
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.9907		Pr(|T| > |t|) = 0.0186			Pr(T > t) = 0.0093	

Table 22: average amount of clinical projects from the publications by publication type sorted by research type (missing data dropped)

Collaborative research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic publications	444	1.653153	.0728052	1.534101	1.510067	1.79624
Applied publications	79	1.405063	.162661	1.445763	1.08123	1.728897
Total	523	1.615679	.0665675	1.522344	1.484906	1.746452
Difference		.24080899	.1782111		-.1050267	.6012064
t = 1.3921						
Satterthwaite's degrees of freedom = 111.594						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9167		Pr(|T| > |t|) = 0.1667		Pr(T > t) = 0.08333		

Independent research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic publications	129	1.689922	.1360425	1.545146	1.420739	1.959106
Applied publications	9	1.888889	.7718024	2.315407	.1091093	3.668669
Total	138	1.702899	.1358314	1.595658	1.434301	1.971496
Difference		-.1989664	.7837006		-1.987697	1.589764
t = -.02539						
Satterthwaite's degrees of freedom =8.50432						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.4028		Pr(|T| > |t|) = 0.8056		Pr(T > t) = 0.5972		

Table 23: average number of citations of patents referring to the publication by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	15.40196	.2653098	19.96886	14.88185	15.92207
Independent publications	1160	16.75969	.6569416	22.37462	15.47076	18.04861
Total	6825	15.63272	.246962	20.40241	15.1486	16.11685
Difference		-1.357726	.7084924		-2.747424	.0319724
t = -1.9164						
Satterthwaite's degrees of freedom = 1559.41						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0278		Pr(|T| > |t|) = 0.0555		Pr(T > t) = 0.9722		

Table 24: Average novelty in knowledge sources of patents referring to the publications by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	4.392129	.2435757	18.33302	3.914627	4.869631
Independent publications	1160	3.568793	.379414	12.92238	2.824378	4.313209
Total	6825	4.252192	.2122352	17.5335	3.836145	4.668239
Difference		.8233355	.4508704		-.0608335	1.707504
t = 1.8261						
Satterthwaite's degrees of freedom = 2233.56						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9660		Pr(|T| > |t|) = 0.0680		Pr(T > t) = 0.0340		

Table 25: average number of citations of patents referring to the publication by publication type sorted by research type

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5079	15.16343	.2752374	19.61537	14.62385	15.70302
Independent publications	1103	16.75708	.6673498	22.16365	15.44766	18.0665
Total	6182	15.44777	.2556556	20.1011	14.9466	15.94895
Difference		-1.59365	.7218804		-3.009652	-.1776439
						t = -2.2076
						Satterthwaite's degrees of freedom =1499.37
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.0137		Pr(|T| > |t|) = 0.0274			Pr(T > t) = 0.9863	

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	586	17.46933	.9384841	22.71829	15.62612	19.31254
Independent publications	57	16.81007	3.490511	26.35278	9.817733	23.8024
Total	643	17.41089	.9086666	23.04146	15.62657	19.19521
Difference		.6592658	3.614474		-6.560704	7.879236
						t = 0.1824
						Satterthwaite's degrees of freedom = 64.3569
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.5721		Pr(|T| > |t|) = 0.8558			Pr(T > t) = 0.4279	

Table 26: average number of citations of patents referring to the publication by research type sorted by research type

Collaborative research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic publications	5079	15.16343	.2752374	19.61537	14.62385	15.70302
Applied publications	586	17.46933	.9384841	22.71829	15.62612	19.31254
Total	5665	15.40196	.2653098	19.96886	14.88185	15.92207
Difference		-2.305899	.9780123		-4.226139	-.3856586
t = -2.3577						
Satterthwaite's degrees of freedom = 689.375						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0093		Pr(|T| > |t|) = 0.0187		Pr(T > t) = 0.9907		

Independent research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic publications	1103	16.75708	.6673498	2216365	15.44766	18.0665
Applied publications	57	16.81007	3.490511	26.35278	9.817733	23.8024
Total	1160	16.75969	.6569416	22.37462	15.47076	18.04861
Difference		-.0529834	3.553734		-7.161109	7.055142
t = -0.0149						
Satterthwaite's degrees of freedom = 60.1647						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.4941		Pr(|T| > |t|) = 0.9882		Pr(T > t) = 0.5059		

Table 27: maximum number of citations of patents referring to the publication by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	23.56884	.4471104	33.6523	22.59234	24.34535
Independent publications	1160	26.59569	1.126155	38.35544	24.238616	28.80522
Total	6825	24.00029	.4177753	34.531391	23.18132	24.81926
Difference		-3.126846	1.211665		-5.503527	-.7501647
t = -2.5806						
Satterthwaite's degrees of freedom = 1545.32						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0050		Pr(|T| > |t|) = 0.0100		Pr(T > t) = 0.9950		

Table 28: average novelty of patents referring to the publication by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	.2086546	.0134909	1.015407	.1822073	.2351019
Independent publications	1160	.2408274	.0386176	1.315267	.1650592	.3165956
Total	6825	.2141228	.0129792	1.072262	.1886794	.2395662
Difference		-.0321728	.0409062		-.1124143	.0480687
t = -.0.7865						
Satterthwaite's degrees of freedom = 1454.72						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.2159		Pr(|T| > |t|) = 0.4317		Pr(T > t) = 0.7841		

Table 29: average novelty of the patents referring to the publication by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	4.392129	.2435757	18.33302	3.914627	4.869631
Independent publications	1160	3.568793	.379414	12.92238	2.824378	4.313209
Total	6825	4.252192	.2122352	17.5335	3.836145	4.668239
Difference		.8233355	.4508704		-.0608335	1.707504
t = 1.8261						
Satterthwaite's degrees of freedom = 2233.56						

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.9660	Pr(|T| > |t|) = 0.0680	Pr(T > t) = 0.0340

Table 30: Average novelty in further development of patents referring to the publication by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5665	1.196602	.1140949	8.587486	.9729321	1.420271
Independent publications	1160	1.207126	.1619914	5.517229	.8892967	1.524955
Total	6825	1.19839	.0986199	8.147338	1.005065	1.391716
Difference		-0.105243	.1981385		-.3990589	.3780104
t = -0.0531						
Satterthwaite's degrees of freedom = 2469.76						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.788		Pr(|T| > |t|) = 0.9576		Pr(T > t) = 0.5212		

Table 31: average time between publication and referring patents (in years) by publication type

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5367	-2.538591	.024676	1.807761	-2.586966	-2.490215
Independent publications	1139	-3.43125	.0613813	2.071562	-3.551683	-3.310817
Total	6506	-2.694868	.0233974	1.887227	-2.740734	-2.649001
Difference		.8926594	.0661557		.7628938	1.022425
t = 13.4933						
Satterthwaite's degrees of freedom = 1527.1						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 1.0000		Pr(|T| > |t|) = 0.0000		Pr(T > t) = 0.0000		

Table 32: average time between publication and referring patents (in years) by publication type sorted by research type

Basic research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	5079	-1.625517	.0306885	2.187082	-1.68568	-1.565354
Independent publications	1103	-2.517679	.0704418	2.339475	-2.655894	-2.379464
Total	6182	-1.784698	.0285026	2.241034	-1.840572	-1.728823
Difference		.892162	.0768364		.7414477	1.042877
t = 11.6112						
Satterthwaite's degrees of freedom = 1547.91						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 1.0000		Pr(|T| > |t|) = 0.0000		Pr(T > t) = 0.0000		

Applied research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Collaborative publications	586	-1.749147	0.606885	2.187082	-1.68568	-1.565354
Independent publications	57	-2.964912	.2672201	2.017468	-3.500218	-2.429606
Total	643	-1.856921	.0828103	2.099858	-2.019532	-1.694309
Difference		1.215766	.2806697		.6557106	1.77582
t = 4.3317						
Satterthwaite's degrees of freedom = 68.0849						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 1.0000		Pr(|T| > |t|) = 0.0000		Pr(T > t) = 0.0000		

Table 33: average time between publication and referring patents (in years) by research type sorted by publication type

Collaborative research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic publications	5079	-1.625517	.0306885	2.187082	-1.68568	-1.565354
Applied publications	586	-1.749147	0.606885	2.187082	-1.68568	-1.565354
Total	5665	-1.638305	.0289134	2.176202	-1.694987	-1.81624
Difference		.1236299	.0911628		-.0553376	.3025974
t = 1.3651						
Satterthwaite's degrees of freedom = 742.691						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9123		Pr(|T| > |t|) = 0.1755		Pr(T > t) = 0.0877		

Independent research publications

Two sample t test with unequal variances

Group	Observations	Means	Std. Err.	Std. Dev.	95% Confidence interval	
Basic publications	1103	-2.517679	.0704418	2.339475	-2.655894	-2.379464
Applied publications	57	-2.964912	.2672201	2.017468	-3.500218	-2.429606
Total	1160	-2.539655	.0682919	2.325938	-2.673645	-2.405666
Difference		.4472332	.2763488		-.1048307	.9992971
t = 1.6184						
Satterthwaite's degrees of freedom = 64.0376						
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9448		Pr(|T| > |t|) = 0.1105		Pr(T > t) = 0.0552		

Table 34: descriptive statistics for the variables in the regression analyses

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
Research type (0: collaborative publication, 1: independent publication)	59024	.760572	.426738	0	1
Government support for publication (0: no support, 1:support)	59024	.4083763	.4915376	0	1
Number of authors	59024	6.423133	4.038367	1	40
Publication year	59024	1998.616	2.494464	1994	2002
Region of the headquarter of the publishing company (1: EU, 2: IL, 3: JP, 4: US)	59024	2.597774	1.398297	1	4
Amount of research publications per million dollars in R&D budget	59024	.2550525	.2809557	0	9.873061
Amount of published clinical projects per million dollars in R&D budget	59024	.0001189	.001091	0	.1109986
R&D budget in million dollar	59024	1463.348	1141.589	0	4846.997
Independent basic research publication	59024	.1075495	.3098132	0	1
Collaborative basic research publication	59024	.6530225	.4760126	0	1
Independent applied research publication	59024	.0178063	.1322481	0	1
Collaborative applied research publication	59024	.2216217	.4153414	0	1

Table 35: probit regression of the effects of the independent variables on the publication type

Probit regression Observation = 46066

Log Likelihood = -11520.832

Independent publication	**Coefficient**	**Standard error**	**Z**	**P>|Z|**	**95% confidence interval**	
Basic research publication	.3321914	.0309232	10.74	0.000	.271583	.327998
Government support received	-1.420486	.0296727	-47.87	0.000	-1.478643	-1.362328
Number of authors	-.1107227	.0031091	-35.61	0.000	-.1168163	-.104629
Publication year	-.29797558	.0057312	-48.81	0.000	-.2909888	-.2685228
Region of the headquarter of the publishing company						
Israel	-.2220258	.5853733	-0.38	0.704	-1.369336	.952847
Japan	-.2376503	.3424659	-0.69	0.588	-.9088712	.4335706
US	.2002977	.3367215	0.59	0.552	-.4596643	.8602598
Amount of research publications per million dollars in R&D budget	.1672929	.0588615	2.84	0.004	.0519265	.2826593
Amount of published clinical projects per million dollars in R&D budget	26.68516	14.2268	1.87	0.061	-1.255454	54.62577
R&D budget in million dollar	0.0001301	.0000174	7.48	0.000	.000096	.001643

Table 36: Negative binomial regression on the amount of clinical projects referring to the publications

Negative binomial regression Observation = 59024

Log Likelihood = -4915.0812

Number of clinical projects referring to a publication	**Coefficient**	**Standard error**	**Z**	**P>|Z|**	**95% confidence interval**	
Basic research publication	.7952864	.1747259	4.55	0.000	.4528298	1.137743
Independent research publication	.5086148	.1433037	3.55	0.000	.2277446	.789485
Number of authors	.1066851	.0100558	10.61	0.000	.0869761	.1263942
Publication year	-.1283203	.0289111	-.4.44	0.000	-.184985	-.0716556
Region of the headquarter of the publishing company						
Israel	.8277134	1.05986	0.78	0.435	-1.249574	2.905
Japan	-.7200563	.165809	-4.34	0.000	-1.045036	-.3950766
US	-.1764241	.1248687	-1.41	0.158	-.4211622	.0683141
Amount of research publications per million dollars in R&D budget	.1679165	.14901	1.14	0.253	-.1200041	.4558371
Amount of published clinical projects per million dollars in R&D budget	224.4235	102.1455	2.20	0.028	24.22195	424.6251
R&D budget in million dollar	-.0000789	.0000566	-1.39	0.163	-.0001898	.0000321

Table 37: Linear regression on the average number of citations of patents referring to the publications

Linear regression Observation = 6825

Average citations of patents referring to the publication	**Coefficient**	**Standard error**	**Z**	**P>|Z|**	**95% confidence interval**	
Independent research publication	.4599579	.7667925	0.60	0.549	-.043195	1.963111
Basic research publication	-2.289048	.952763	-2.40	0.016	-4.156761	-.4213355
Received government support	1.276287	.5268689	2.42	0.015	-2.309114	-.243459
Number of authors	-.0620532	.0400134	-1.55	0.121	-.140492	.0163857
Publication year	-3.165921	.1305978	-2.42	0.015	-.5726045	-.0605798
Region of the headquarter of the publishing company						
Israel	-10.692926	3.540015	-3.00	0.003	-17.56879	-3.689724
Japan	-3.98519	.9133305	-4.36	0.000	-5.775603	-2.194777
US	-2.737727	.5508643	-4.97	0.000	-3.817593	-1.657861
Amount of research publications per million dollars in R&D budget	-1.078801	.9575457	-1.13	0.260	-2.95589	.79822873
Amount of published clinical projects per million dollars in R&D budget	-159.5936	59.82463	-2.67	0.008	-276.8685	.-42.31861
R&D budget in million dollar	-.0003119	.00002944	-1.06	0.289	-.000889	.0002652

Table 38: Linear regression on the average number of citations of patents referring to the publications sorted by both research type and publication type

Linear regression Observation = 6825

Average novelty of sources used in patents referring to the publications	**Coefficient**	**Standard error**	**Z**	**P>|Z|**	**95% confidence interval**	
Independent basic research publication	-2.919009	1.152312	-2.53	0.011	-5.1779	-.6601175
Collaborative basic research publication	-2.214392	1.102085	-2.01	0.045	-4.374823	-.0539601
Independent applied research publication	.0682909	3.229649	0.02	0.983	-6.262829	6.399411
Received government support	-.6676631	.4479419	-1.49	0.136	-.1545769	.210443
Number of authors	-.0404611	.0379129	-1.07	0.286	-.1147821	.03386
Publication year	.1892414	.127513	1.48	0.138	-.0607239	.4392066
Region of the headquarter of the publishing company						
Israel	-4.219345	.782287	-5.39	0.000	-5.752871	-2.685818
Japan	-.58222592	.8419704	-0.69	0.489	-2.232784	1.0682566
US	-.3061352	.4778494	-0.64	0.522	-1.242869	.6305989
Amount of research publications per million dollars in R&D budget	-.1249685	.8961327	-0.14	0.889	-1.881668	1.631732
Amount of published clinical projects per million dollars in R&D budget	-79.70524	20.31679	-3.92	0.000	-119.5325	-39.87799
R&D budget in million dollar	-.00001383	.0002441	-0.57	0.571	-.0006168	.0003401