INNOVATION IN THE BIOPHARMACEUTICAL INDUSTRY

A Dissertation
Presented to
The Academic Faculty

By

Vitor Kamada

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Economics

Georgia Institute of Technology

August 2017

INNOVATION IN THE BIOPHARMACEUTICAL INDUSTRY

Approved by:

Dr. Usha Nair-Reichert, Advisor
School of Economics
Georgia Institute of Technology

Dr. Mikhail Klimenko
School of Economics
Georgia Institute of Technology

Dr. Levent Kutlu
School of Economics
Georgia Institute of Technology

Dr. Richard Barke
School of Public Policy
Georgia Institute of Technology

Dr. Patrick S. McCarthy
School of Economics
Georgia Institute of Technology

Date Approved: April 26, 2017

ACKNOWLEDGMENTS

Several people contribute to this journey. Any errors are my own.

A very special gratitude goes out to my advisor Prof. Usha Nair-Reichert. I could not have imagined having a more careful advisor. I am also grateful to the Econometricians: Dr. Levent Kutlu and Dr. Patrick McCarthy. Thanks for the hard technical questions which forced me to rethink my empirical strategies in other ways. I would like to thank Prof. Mikhail Klimenko for insightful comments about clinical trials. I would like to thank Prof. Richard Barke for bringing practical aspects of technological alliances.

My sincere thanks also goes to Prof. Vivek Ghosal, Prof. Marco Ceccagnoli, and Prof. Henry Sauermann. I learned so much about Innovation in their courses. I am also grateful to Prof. Matthew Higgins to let me use the BioSciDB and Pharmaprojects database. I will forever be thankful to the Chief Scientist Dennis Folds and Director of Business Strategy Marty Broadwell for the opportunity of working with economic analysis of R&D projects at Georgia Tech Research Insitute (GTRI).

I thank Prof. Marco Ceccagnoli, Prof. Jonathan Clarke, and Prof. Peter Thompson for letting me teaching Statistics at Georgia Tech College of Business. I am also grateful to Prof. Tatiana Rudchenko, who shared her experience in teaching Statistics. I also thank Prof. Vivek Ghosal for advising and letting me teaching Macroeconomics at Georgia Tech School of Economics.

I would like to thank Prof. Byung-Cheol Kim for the great experience of being his teaching assistant (TA). I really loved being his TA for the PhD courses: Mathematics for Economists, Industrial Organization I, and Microeconomic Theory II. I am also grateful to Prof. Parks Dodd, who helped me with Forecasting, Time Series, and Eviews. I enjoyed very much being his TA for Econometric Modeling, and Business Forecasting. I also worked as TA for Prof. Shatakshee

Dhongde and Prof. Christine Ries. I would like to thank both for inciting me to reflect about course management/practice at undergraduate level.

I thank Prof. David Laband and Prof. Tibor Besedes for their stronger support during storm periods.

I would like to thank all my Professors from Georgia Tech: Prof. Haizheng Li, Prof. Jerry Thursby, Prof. Juan Moreno-Cruz, Prof. Matthew Oliver, Prof. Erik Johnson, Prof. Johnson Kakeu, Prof. Ruth Uwaifo Oyelere, and Prof. Jason Borenstein.

I am grateful to Rachel Goddard who missed some Salsa Parties to read the manuscript. I am grateful to Prof. Beryl Martinson who also helped with my English.

I would like to thank Tony Gallego, Jyldyz Hughes, and Sina Mehdikarimi for the great administrative support.

Finally, I would like to thank my classmates: Chris, Yeong, Vijay, Jingwen, Mayank, Carrie, and Tingting.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

AACT	Aggregate Analysis of ClinicalTrials.gov
ATC	Anatomical Therapeutic Chemical
ATET	Average Treatment Effect on the Treated
CROs	Contract Research Organizations
CTTI	Clinical Trials Transformation Initiative
DMC	Data Monitoring Committee
EPO	European Patent Office
FD&C	Federal Food, Drug, and Cosmetic
FDA	US Food and Drug Administration
FDAAA	Food and Drug Administration Amendments Act
FIML	Full Information Maximum Likelihood
FOIA	Freedom of Information Act
ICMJE	International Committee of Medical Journal Editors
IRB	Institutional Review Board
JPO	Japan Patent Office
LIML	Limited Information Maximum Likelihood
M&A	Mergers and Acquisitions

MeSH	Medical Subject Headings
mRNA	Messenger Ribonucleic Acid
NCE	New Chemical Entity
NIH	National Institutes of Health
NLM	U.S. National Library of Medicine
NNME	Nearest-Neighbor Matching Estimator
NPOs	Nonprofit Organizations
OECD	Organisation for Economic Co-operation and Development
PSM	Propensity Score Matching
R&D	Research and Development
RECAP	Recombinant Capital
SDC	Securities Data Company
USPTO	United States Patent and Trademark Office
WHO	World Health Organization

SUMMARY

Three aspects of the drug development process are little understood in the literature: 1) which type of organization has comparative advantage in conducting clinical trials, firms or nonprofit organization (NPOs)?; 2) what determines the timing of technological collaboration, a step usually far away from drug commercialization?; and 3) how the strategies making, allying, and buying R&D impact the probability of drug approval? Using data on over 20,000 clinical trials, 800 technological alliances, and 14,000 R&D projects, the main findings of this econometric study are: 1) firms complete each phase of clinical trial faster than nonprofit organizations; 2) licensor scientific and technological specialization is the main factor to speed up the technological collaboration; and 3) R&D projects developed by research alliance have higher probability of drug approval than internal R&D projects, but R&D projects originated from M&A have lower probability of drug approval than internal R&D projects. Together these findings inform policy makers and managers on how to foster the innovation performance in the biopharmaceutical industry.

CHAPTER 1

INTRODUCTION

Innovation has a high impact on life span and social welfare. In the biopharmaceutical industry, innovation is mainly represented by a new drug; whereas a patent is an important measure of invention. There is long shot between the invention, the discovery of a chemical or biological compound, until the drug commercialization. Before drug approval, the compound must be tested in terms of safety and efficacy under the clinical trial phases I, II, and III. Three aspects of the drug development process are little understood in the literature: 1) which type of organization has comparative advantage in conducting clinical trials, firms or nonprofit organization (NPOs)?; 2) what determines the timing of technological collaboration or research alliance?; and 3) how the strategies making, allying, and buying research and development (R&D) impact the probability of drug approval? This dissertation supports the thesis that: 1) firms complete clinical trials faster than NPOs, an evidence of relative efficiency of firms; 2) licensor scientific and technological specialization is the main factor to speed up the technological collaboration; and 3) considering internal R&D projects as benchmark, R&D projects developed by research alliance have higher probability of drug approval, but R&D projects originated from mergers and acquisitions (M&A) have lower probability of drug approval.

In chapter 2, using data on over 20,000 clinical trials started during 2006 to 2011, this study compares the duration of phase I, II, III, and IV between firms and nonprofit organizations (NPOs). After controlling for clinical trial characteristics, disease variation, external collaboration, scale of the organization, and other factors, we found that firms complete each phase of clinical trial on average 13 months early than NPOs. We interpret this result as evidence that firms are more efficient in late-stage of R&D process in comparison to NPOs. We also found that the external support by a

NPO delays the duration of clinical trials on average 4 months; whereas the external support by a firm doesn't affect the duration of clinical trials. Size of the organization, and previous experience in clinical trials don't appear to affect the duration of clinical trials. Together our results inform policymakers on the debate of publicization vs privatization of clinical trials.

In chapter 3, using a sample on over 800 technological alliances in the biopharmaceutical sector, we investigate how scientific and technological specialization affect the timing of technological collaboration. We propose a new measure of specialization and timing of technological collaboration. As each alliance targets a specific technology (for example: monoclonal antibodies, recombinant DNA, and stem cells) or disease (for example: leukemia, diabetes, and Parkinson's disease), we could count the scientific publications and patents that have these keywords in the title or abstract. Related to the timing of technological collaboration, we measure the timing between the licensor first scientific publication (that cites a specific technology or disease) and the starting date of the alliance. After controlling for several factors at alliance and firm level, we found that licensor specialization is the main factor to speed up the technological collaboration. An increase of 10% in the scientific specialization speeds up the technological collaboration on average 4 months, and an increase of 10% in the technological specialization speeds up the technological collaboration on average 5 months. Furthermore, our results suggest that licensee scientific and technological specialization in the same technology or disease of the licensor don't affect the timing of technological collaboration. A finding that is aligned with the theory of comparative advantage, as specialization must be in different tasks or stages, in order to explore the differences in opportunity cost and potential complementarities between activities.

Chapter 4 compares the innovation productivity of the strategies: making, allying, and buying. Using data on over 14,000 R&D projects started during 1989 to 2007 in the pharmaceutical industry, we found that if external R&D projects are defined as projects developed in research alliance and projects originated from M&A, there is no difference between the probability of drug approval of internal R&D projects and external R&D projects. However, when the external R&D projects is split in a relevant category research alliance and M&A, R&D projects developed by research alliance have higher probability of drug approval than internal R&D projects, but R&D projects originated from M&A have lower probability of drug approval than internal R&D projects. We also found that the success of a research alliance or M&A depends positively on who is the partner or

the targeted firm. Research alliances between big pharmaceutical firms have higher probability of success than research alliances between a big pharmaceutical firm and small/medium firm. Furthermore, the results suggest that despite the firms strategically use external R&D to strengthen the R&D portfolio in a specific anatomical therapeutic class or diversify to different anatomical therapeutic class, these strategies are not predictors of innovation performance in terms of drug approval. Together these findings suggest that big pharmaceutical firms are not in-licensing "lemons" projects or at least complementarities/synergies generated by the partnership more than compensated loss due transaction cost and asymmetric information problems.

Chapter 5 presents the conclusion, limitations, and suggestions for future research.

Next section summarizes the evolution of drug discovery process, a background reading for chapters 2, 3, and 4.

1.1 The Evolution of Drug Discovery Process

In this section, we describe the evolution of drug discovery process from random screening through rational design to genetic engineering, highlighting the consequences on the market structure of biopharmaceutical sector.

Random screening is trial and error method with little scientific guidance to find a drug. Without the understanding of the causes of the disease or the mechanism of action of the drug, it is possible to infer that a drug works by randomized clinical trial. Until 1950s, most part of chemical drugs were discovered by Swiss and German firms, using the random screening method [Schwartzman (1976), Gambardella (1995)].

The birth of pharmaceutical industry can be traced to the Swiss and German chemical firms like Bayer, Ciba[1], Hoechst[2], and Sandoz[3]. In the late 19th century, these firms were already the world leader producers of organic chemical compounds like dyestuffs, a substance originally used to add a color. Closer connected with academic institutions, they were also the first firms to create industrial laboratories to investigate the therapeutic effects of dyestuffs [Chandler (2009), Scherer (2010), Malerba and Orsenigo (2015)]. The acetylsalicylic acid, known as Aspirin, is an example

[1] In 1971, Ciba merged with Geigy to form Ciba-Geigy. In 1996, Ciba-Geigy merged with Sandoz to form Novartis.
[2] In 1999, Hoechst merged with Rhône-Poulenc to form Aventis. In 2005, Aventis became a subsidiary of Sanofi.
[3] In 1996, Sandoz merged with Ciba-Geigy to form Novartis.

of one dyestuff formulation tested in the Bayer laboratory. Other example of dyestuff variant (coal-tar) is the class of sulfa drugs discovered by IG Farben[4] in 1935. Before the mass production of penicillin in 1944, sulfa drugs were the only effective antibiotic available to treat diseases such as meningitis, pneumonia, and gonorrhea.

Until the World War II, Swiss and German firms were the drug innovators; whereas the big pharmaceutical firms of other countries were the national distributors of synthetic coal-tar drugs. In the random screening era, Swiss and German firms run large scale experiments, tested millions of compounds, and accumulated extensive libraries about the effects of these compounds. The economies of scope and scale in R&D process were working as barriers to entry [Henderson and Cockburn (1996), Chandler (2009), Malerba and Orsenigo (2015)]. However, in the interwar period American and other European firms started to develop research capabilities, as German firms lost patents, facilities, and access to global markets. For example, during World War I, US government sequestered Merck & Co (US subsidiary) from Merck Group (headquarter in Darmstadt, Germany). Ironically, Merck & Co will become later the most successful firm in the rational drug design and the genetic engineering era.

In the rational drug design process, biochemistry, microbiology, and enzymology provide guidance about the ideal molecule that would eliminate the pathological agent [Gambardella (1995), Scherer (2010)]. The knowledge of a biological or chemical target restricts considerably the screening process. X-ray crystallography, nuclear magnetic resonance, and sophisticated mathematical and computational models help scientists to design the molecular structure that is likely to have the desirable therapeutic effects.

The classical example of rational design is the ulcer drug, Tagamet from GlaxoSmithKline. In the 1970s, scientists knew that histamine stimulates the production of stomach acid. The problem was to find the histamine receptor antagonist to suppress the gastric acid. Based on the structure of histamine, the scientists of GlaxoSmithKline designed and tested about 700 compounds during 10 years before discovering the H2-antagonist Tagamet [Azoulay (2002), Scherer (2010)]. Another famous example is the antidepressant, Prozac, a selective serotonin reuptake inhibitor (SSRI) discovered by Eli Lilly in 1972. Scientists knew that if serotonin, a chemical secreted by the brain, is absorbed too fast by the nerve cells, depression is triggered. Then, Eli Lilly's researchers started

[4]In 1951, IG Farben was split and absorbed by BASF, Bayer, Hoechst, and other firms.

to design and screen chemical compounds that slow down the absorption of serotonin in a selective way to minimize side-effects. Prozac was the first blockbuster of Eli Lilly, partly because it presented less side-effects than non-serotonin treatments for depression [Gambardella (1995)].

The next technical revolution is the drug discovery by using genetic engineering. In the 1970s, the advent of biotechnology made the drug discovery process more science-based and dependent of academic institutions [Gambardella (1992), Zucker et al. (2002), Scherer (2010)]. The discovery of recombinant DNA (rDNA) and monoclonal antibody (mAb) create unparalleled possibilities to diagnose and treat diseases. The recombinant DNA technique, the transference of DNA segment from one organism to another organism, was developed in Stanford and University of California (San Francisco) in 1973. Important application of the technique includes identification of genes function and production of substances like recombinant insulin and growth hormone. The monoclonal antibody (mAb) technique, clones of a unique parent cell, was developed by University of Cambridge (UK) in 1975. This "magic bullet", in the sense that selectively targets a substance or organism, allowed a more effective treatment of several types of cancer.

It is worth mentioning that government sponsored organizations always have an important role in drug discovery, especially in biological compounds like vaccine [Trusheim et al. (2010)]. Top research centers in microbiology and immunology like Pasteur Institute in Paris and Koch Institute in Berlin conferred competitive advantage in vaccines respectively to French firm Sanofi and German firms. Furthermore, in a sample of Top 19 drugs in terms of therapeutic impact introduced between 1965 to 1992, Cockburn and Henderson (2000) found that: 1) 14 of these drugs were discovered with essential inputs from public sector; and 2) among the 5 drugs that were discovered without essential inputs from public sector, 4 drugs were discovered by random screening process.

Other consequence of the biotechnology revolution is the obsolescence of the "chemical capabilities", that allowed the entrance of small specialized firms in the drug discovery process. Without the organizational rigidity and bias toward the old technology, biotech firms started to explore the new technology opportunities, serving as broker between academic institutions and the consolidated pharmaceutical firms [Zucker et al. (2002), Stuart et al. (2007)]. Among thousands of small specialized firms, only two biotech firms become successful vertically integrated firms Amgen and Genzyme [Chandler (2009)]. The incumbents weren't replaced, because long, costly, risky clinical trials and distribution network remain barriers to entry [DiMasi et al. (2003), DiMasi and

Grabowski (2007), Rothaermel (2001)]. Other reason is that several consolidated pharmaceutical firms also moved fast to integrate the new technologies via research alliances, joint venture, and M&A [Gambardella (1992), Rothaermel and Deeds (2004), Mittra (2007), Malerba and Orsenigo (2015)].

In 1980s, German firms, the old leaders, were technologically behind to American firms like Merck, Pfizer, Eli Lilly, and biotechs. Chandler (2009), Malerba and Orsenigo (2015) speculate that the main cause was that German academic institutions fell behind in comparison to American academic institutions. Outside of United States, only Cambridge (UK) rivals to the biotech hubs in San Francisco Bay and Boston. However, the problem of German firms is not about geographic spillover, but about strategy. Swiss firms like Novartis and Hoffmann-La Roche developed internally and acquired externally research capabilities via research alliances, joint venture, and M&A. For example, Genentech, the first biotech firm and founded by Herbert Boyer, the creator of recombinant DNA technique, was acquired by Hoffmann-La Roche in 1990. Another example is Chiron the main competitor of Genentech. Ciba-Geigy (in 1996, this Swiss firm and Sandoz formed Novartis) started a joint venture with Chiron in 1986 and acquired Chiron 8 years late [Chandler (2009)].

Therefore, we showed how the evolution of drug discovery process shaped the complex landscape of biopharmaceutical sector, creating technological opportunities for new entrants and followers to challenge the incumbents.

CHAPTER 2

DURATION OF CLINICAL TRIALS: A COMPARISON BETWEEN FIRMS AND NONPROFIT ORGANIZATIONS

2.1 Introduction

It is a long process to bring new drugs to the market. Drug development can be divided in three main stages: discovery of the chemical or biological compound, preclinical studies with animals, and clinical trials with human beings. The last stage can be divided in four phases: 1) Phase I - small scale study that tests the safety of the compound; 2) Phase II - small scale study that checks the efficacy of the compound, and generates enough information to kill the project or to advance it to further development; 3) Phase III - large scale confirmatory study about the safety and effectiveness of the compound submitted to the regulatory authority for drug approval; and 4) Phase IV - regulatory authority might require postmarket information about the safety and effectiveness the compound.

Scholars often conceive drug development as a patent race where winner takes all [Kamien and Schwartz (1974), Fudenberg et al. (1983), Denicolo (1996)]. However, obtaining a patent is not the end of the race, but the beginning given the higher attrition rate of each phase of clinical trial. Neither drug development is a winner takes all game, a patent blocks rivals to use the same chemical compound to treat a specific condition, but it doesn't impede other firms to develop superior treatments with less side effects based on other chemical or biologic compound. Lichtenberg and Philipson (2002) found that between-patent competition (new patents from brand name firms) is as fiercely as competition within patents (generic drugs). DiMasi and Paquette (2004) studied the development history of several drugs and found that almost always at least one follow-on drug

("me-too") was in the clinical trial before the breakthrough drug ("first in class") was approved in United States from 1960 to 1998, and that in several cases, the first drug in class was not the first to start the clinical trial.

Big pharmaceutical firms even buy expensive priority review vouchers to accelerate the drug approval process. Under Section 529 to the Federal Food, Drug, and Cosmetic Act (FD&C Act), US Food and Drug Administration can award priority review vouchers to firms who received a drug approval for a rare pediatric disease. The voucher reduces the FDA decision from 10 months to 6 months. In August 2015, AbbVie payed $350 million for a voucher from United Therapeutics. The first voucher issued was bought for $67.5 million by Sanofi from BioMarin Pharmaceutical in 2014. Sanofi, in alliance with Regeneron, were the first to show that PSCK9 inhibitor lowers the cholesterol. However, Amgen run clinical trials faster for its own PSCK9 drug (Repatha). Whereas Amgen tracked patients during 12 weeks; Sanofi tracked patients during 24 weeks. Although, Amgen was the first to finish the clinical trials and to apply for drug approval, Sanofi received first the drug approval for Praluent, because of the priority review voucher [Winslow and Walker (2014), Loftus (2015)].

Although the duration of clinical trial is a critical element of drug development race, literature focuses on cost estimation [DiMasi et al. (1991), DiMasi (2001), DiMasi et al. (2003), DiMasi and Paquette (2004), Adams and Brantner (2006), DiMasi and Grabowski (2007), Berndt and Cockburn (2014)] and probability of success of clinical trials [Danzon et al. (2005), Guedj (2005), Arora et al. (2009)]. Related to the duration of clinical trials, the studies [DiMasi et al. (1991, 2003), Guedj (2005), DiMasi and Grabowski (2007)] restrict themselves to present descriptive statistics about the length of phase I, II, and III. The literature also ignores the importance of nonprofit organizations (NPOs) and phase IV in the clinical trials. However, our data about clinical trials started during 2006 to 2011 indicate that: 1) the principal sponsor of 43% of clinical trials were a university, hospital, federal agency, or other NPO; 2) among the clinical trials developed with the support of other organizations, almost half (47%) receives a support of a NPO; and 3) phase IV represents 25% of clinical trials conducted by NPOs, but only 10% of clinical trials conducted by firms.

This paper aims to fill the gaps in the literature mentioned above. We support the thesis that firms complete clinical trials in shorter time than NPOs. After controlling for several characteristics of clinical trials, our OLS baseline estimation indicates that firms complete clinical trials 13, 14,

16, and 11 months early than NPOs for respectively phases I, II, III, and IV. We also found that external support by firms doesn't affect the duration of clinical trials, but the NPOs support delays clinical trials in 2, 5, 6, and 4 months for respectively phases I, II, III, and IV. The results are robust for Nearest-Neighbor Matching Estimator (NNME) and for different specifications that control for unobservable quality of clinical trials (presence or not of data monitoring committee) or disease variations (363 subclasses).

Our findings support the theoretical prediction of Aghion et al. (2008) that firms have comparative advantage in the late-stage research (clinical trials). The intuition is that creative freedom provided by NPO environment matters more in the early-stage research (discovery phase); but in the late-stage research (clinical trials), control and direction towards higher economic payoff is more important. Our findings also contribute to the debate of publicization vs privatization of clinical trials. Several scholars [Lewis et al. (2007), Baker (2008), Jayadev and Stiglitz (2009)] defend that clinical trials should be considered as a public good, and consequently public funded. Lewis et al. (2007) propose:

> " *... to establish an independent testing agency to conduct clinical trials at a national testing facility, under specified conditions of transparency. Drug companies would no longer directly compensate scientists for evaluating their own products; instead, scientists would work for the testing agency, which would be supported by funds collected from taxes upon the pharmaceutical industry and/or from general tax revenue*".

Any policy has different dimensions and trade-offs to be evaluated. If public funded means conducting the clinical trial under a NPOs structure, our study points out for a loss of efficiency in one dimension: completion timing.

2.2 Literature Review

This literature review aims to summarize the theoretical foundations for our thesis that firms have comparative advantage in performing clinical trials in comparison to NPO. We also present a short summary of empirical findings related to the duration and other aspects of clinical trials.

2.2.1 Theoretical Foundations

Several scholars [Nelson (1959), Merton (1973), Partha and David (1994), Polanyi (2000), Sauermann and Stephan (2013)] have been pointed out that NPOs and firms are organizations with different institutional logics, reward systems, and goals. Specifically, Aghion et al. (2008) modelled the advantage and disadvantage between research conducted by NPO vs Firm based on the property rights literature [Grossman and Hart (1986), Hart and Moore (1990), Aghion and Tirole (1994), Hart (2003)]. Traditionally, the literature justified research conducted by NPO based on spillovers and appropriability problems of basic research [Nelson (1959), Arrow (1962)].

However, Aghion et al. (2008) defends that creative freedom is the main factor to justify the comparative advantage of NPO in the early-stage research. The logic is that freedom of pursing multiple paths of interest is more important in the early-stage research; whereas focus on higher economic payoff matters more in the late-stage research. In other words, freedom fosters the creativity of scientists, and consequently, the probability of invention, but the conversion of invention into innovation depends more on targeted effort towards specific goals. Other argument is that in the early-stage research, academic freedom decreases the labor cost normalized by the motivation and effort of scientists; but in the late-stage research, labor cost is a marginal factor compared to the urgency of finishing the R&D project and realize the potential payoff.

Different from Aghion et al. (2008) that conceive a linear R&D process with distinct and divisible phases, Lacetera (2009) raises the question if a whole R&D project should be performed inside the firm or outsourced to NPO. Lacetera (2009) argues that outsourcing a R&D project to a NPO is a credible commitment device that a firm will respect the scientific values of the R&D project without changing it based on economic considerations about profitability. Lacetera (2009) predicts that long-term R&D project with higher level of uncertainty is more likely to be performed by a firm; whereas R&D project with broader applicability is more likely to be outsourced to NPO. The logic is that if the uncertainty is high, firms want to control the research agenda and direct it to more certain profitable paths; but if uncertainty is low and the R&D project has broader applicability (higher economic payoff), firms prefer to cede control to NPO.

By the other hand, Budish et al. (2015) defend that firms underinvest in long-term research and prefer short-term research. The logic is that due to corporate short-termism and fixed patent

term, firms has less incentives to invest in R&D projects, in which the lag between patents and commercialization is longer. It is worth noting that if long-term R&D project of Lacetera (2009) is interpreted to include early-stage research of Aghion et al. (2008), both models produce opposite predictions.

2.2.2 Empirical Literature of Clinical Trials

Empirical literature in economics focused mainly on cost estimation [DiMasi et al. (1991), DiMasi (2001), DiMasi et al. (2003), DiMasi and Paquette (2004), Adams and Brantner (2006), DiMasi and Grabowski (2007), Berndt and Cockburn (2014)] and probability of success of clinical trials [Danzon et al. (2005), Guedj (2005), Arora et al. (2009)]. Budish et al. (2015), Isakov et al. (2016) covered other aspects related to clinical trials described in paragraphs below. Concerned to the duration of clinical trials, part of the literature [DiMasi et al. (1991, 2003), Abrantes-Metz et al. (2004), Guedj (2005), DiMasi and Grabowski (2007)] restricts itself to present only descriptive statistics about the length of phase I, II, and III. The exception is Dranove and Meltzer (1994), Getz and De Bruin (2000). Dranove and Meltzer (1994) investigated 574 approved drugs from 1950 to 1986, and found that firms and US Food and Drug Administration (FDA) accelerate the drug development and approval process of important drugs. Getz and De Bruin (2000) interviewed managers of five companies to assess business practices that speed up the development cycle of biopharmaceutical industry.

DiMasi et al. (1991, 2003), Guedj (2005), DiMasi and Grabowski (2007) has been estimating the total capitalized R&D cost of a new drug. The last estimate of DiMasi and Grabowski (2007) for new chemical and biologic molecule is respectively $1.24 and $1.31 billion. Both values include opportunity cost and are adjusted for the failure risk. Although DiMasi and Grabowski (2007) estimate the average development time to compute the opportunity cost, nothing else is explored about the duration of clinical trial. Instead of estimated the cost of drug development, Isakov et al. (2016) estimated the cost of type I error (the cost of approve an ineffective drug) and type II error (the cost of not approve an effective drug) based on Bayesian decision framework. Isakov et al. (2016) concluded that FDA standards deviate from the social optimum, that is, FDA demands larger sample and too conservative threshold of statistical significance for deadly illness, but smaller sample and lower threshold of statistical significance than the optimum for mildest illness.

In a sample of 40 big pharmaceutical firms during 1984-2001, Guedj (2005) investigated the probability of drug candidates to advance to the next phase of clinical trial. Although Guedj (2005) focused on the comparison of drug approval rate between internal vs external R&D projects, they also presented descriptive statistics about the duration of clinical trials. Compared to Guedj (2005), the duration of clinical trials in our Table 2.2 is somewhat shorter. One explanation is that the samples and the period analyzed are different. In fact, our analysis of ClinicalTrials.gov data suggests that clinical trials are getting slightly shorter over time.

Budish et al. (2015) showed that an increase in the patient five-year survival rate decreases the number of cancer clinical trials, specially for the case of privately financed trials. As Budish et al. (2015) argue that long survival rate is related to long commercialization lag and preventive cancer clinical trials, they conclude that firms underinvest in long-term clinical trials (cancer prevention) in comparison to short-term clinical trials (cancer treatment).

In a sample of 574 approved drugs from 1950 to 1986, Dranove and Meltzer (1994) found that the timing between patent application and FDA approval is shorter for important drugs. Dranove and Meltzer (1994) classified the drugs in level of importance based on: 1) citations in medical textbooks, scientific papers, and patents; 2) approval in major markets like United States, Europe, and Japan; and 3) sales of top 75 selling drugs in United States from 1960 to 1990.

In a sample of 109 firms and 725 NCEs (new chemical entity) approved during 1981 to 1999, Getz and De Bruin (2000) calculated that the average drug development time was 12.2 years for the firm bottom quartile, 5.6 years for the firm median quartile, and 2.7 years for the firm top quartile. These figures cannot be interpreted as evidence of higher performance of the firms in the top quartile, because no control was employed, and the comparison was made among firms with different disease portfolio. Getz and De Bruin (2000) pointed out that firms in the top quartile have superior business practice, such as: global project planning, realistic protocols, active collaboration with regulatory authorities, adoption of modern information technology (IT), and project team empowerment.

2.3 Hypothesis

Modelling the trade-off between research developed by firms vs nonprofit organizations (NPOs), Aghion et al. (2008) argue that the innovation process is more efficient, when the early-stage re-

search is conducted by a NPOs and the late-stage research by a firm. The logic of this labor division is that in the early-stage research, the creative freedom supported by NPO environment increases the effort and motivation of scientists and in the same time decreases its wages; whereas in the late-stage research, control and direction towards higher economic payoff is more important.

Biopharmaceutical sector is a perfect place to test the theoretical predictions of Aghion et al. (2008), because R&D process of this industry can be decomposed in distinct phases with clear regulatory milestone. Other characteristic of this industry is that patent application occurs far advanced from drug commercialization usually before the preclinical phase. Inevitably substantial part of patent fix term protection of 20 years in US is lost due to the clinical trial period. Minimize the timing gap between drug discovery and commercialization is critical to preserve patent rent and it is aligned with the goal of profit maximization, but not necessary aligned with scientific ethos [Nelson (1959), Merton (1973), Partha and David (1994), Polanyi (2000), Sauermann and Stephan (2013)].

Labor cost is a critical component of clinical trials. NIH (2015) reported that NIH Clinical Center budget for fiscal year 2014 was $402 millions, and about $251 millions (62% of total budget) was destined to cover the labor cost. Stern (2004) showed that post-Ph.D. biologists accept substantial lower salary in exchange of freedom to pursue their own research interest and ability to publish the results. If by one side, NPOs have labor cost advantage in the early-stage research; by other side, labor cost might be less important in the late-stage research. For example, closer to get a drug blockbuster, the wage of scientists represents a smaller percentage of the expected economic payoff of finishing the R&D project.

Furthermore, several tasks related to the clinical trials are closer to regular business tasks than academic research. Advertising effort in recruiting subjects and data management (monitoring and recording data) are considered less glamorous activity than research at drug discovery phase. Getz and De Bruin (2000), Drennan (2002), Fleischhacker et al. (2015) point out that the main practical problem of clinical trials is to recruit subjects. In some cases, organizations might compete fiercely for patients. The average American often demands monetary compensation for participating in the clinical trials, and are not inclined to give the written consent after being informed about the risks of enrolling in it. NPOs are more likely to be restricted to a specific geographic location; whereas firms have more flexibility to conduct clinical trials in other countries or to rely on CROs (Contract Research Organizations) for recruitment and related tasks [Dranove and Meltzer (1994),

Getz and De Bruin (2000), Mirowski and Van Horn (2005)]. Drennan (2002) affirms that CROs pay the doctors for recruiting patients. According with our descriptive statistics of section 2.4, firms conduct on average more clinical trials in different countries with larger sample than NPOs, suggesting some efficiencies in recruiting process.

Therefore, in comparison to NPOs, firms might be more flexible and have more mechanisms to accelerate projects with higher expected payoff, including paying higher salaries, hiring more scientists, and outsource tasks to CROs. We hypothesize that firms are more efficient in managing clinical trials than NPOs, consequently the duration of clinical trials conducted by firms is shorter than by NPOs:

Hypothesis 1: Ceteris paribus, firms complete clinical trials in shorter time than NPOs.

Azoulay (2003), Mirowski and Van Horn (2005) point out that firms started since 1980s to replace the academic cooperation by CROs, specially for clinical trial tasks. CROs often advertise their services as cheaper than NPOs. Mirowski and Van Horn (2005) enumerate some advantages of CROs such as: custom-made information technologies, capacity to screen a large pool of subjects in different sites, and independent IRB (Institutional Review Board). The National Research Act of 1974 mandates that every organization that receives federal funding must have a IRB to monitor the clinical trials with human beings. Originally, almost all IRB was localized in NPOs, in particular universities. However, the FDA allowed the use of independent IRB in 1981. This explain why the clinical trials conducted by NPOs is more likely to have an external data monitoring committee than clinical trials conducted by firms. Mirowski and Van Horn (2005) argue that the IRB of NPOs views the clinical trial as a small-scale and geographically isolated procedure conducted by a lonely scientist; whereas the independent IRB of CROs has a more practical view of clinical trials in terms of cost and timing. Azoulay (2003) points out that CROs are Taylorist organizations focus on narrow and specific tasks; whereas Pharmaceutical and Biotechs firms are functional flexible organizations with high paid and skilled labor. Azoulay (2003) defends that modern clinical trial is a set of complex tasks, and it is completely inefficient in terms of quality to outsource the coordination of clinical trials to CROs. Azoulay (2003) argues that workers of CROs are poor paid and trained, and unable to capture tacit knowledge or soft data from clinical trials. In fact, firms outsource less important tasks or R&D projects to CROs, and keep the important clinical trials (higher expected

value, for example potential blockbuster) to in-house team.

At first, we could expect that any kind of external collaboration would speed up the clinical trials. However, differently from the discovery phase characterized by an open-ended problem with multiple paths, clinical trials are more like closed-end problems with well-delimited protocol with clear goals. Given these differences, cooperation among organizations in clinical trials might involve duplicity of effort and resources. Even a small organization might have enough scale to run a clinical trial without any external support. Neither theoretically or empirically it is clear, if the benefits of cooperation overcome the cost of cooperation. For example, some authors [Danzon et al. (2005), Arora et al. (2009)] claim that alliance in the discovery phase increases the probability of drug approval; whereas others claim the opposite [Pisano (1997), Guedj (2005)]. By the arguments above, if NPOs are less efficient in conducting clinical trials than firms, probably NPO external support is more likely to delay clinical trials than the firm external support.

Hypothesis 2: Ceteris paribus, relative to firm external support, NPO external support delays the completion of clinical trials.

2.4 Data

The main database is the Aggregate Analysis of ClinicalTrials.gov (AACT) compiled by the Clinical Trials Transformation Initiative (CTTI). The CTTI is a public-private partnership composed by FDA, National Institutes of Health (NIH), pharmaceutical and biotech firms, academic institutions, patient advocacy groups, and other organizations, that aims to promote the "best practices" related to clinical trials. ClinicalTrials.gov is a website maintained by the U.S. National Library of Medicine (NLM) at the NIH. We use the AACT relational database instead of raw data available on ClinicalTrials.gov website, because the AACT database contains extra information about clinical trials not available on ClinicalTrials.gov website, such as: classification of principal sponsor and collaborators of clinical trials, secondary location of clinical trials, presence or not of (DMC), description about potential conflict of interest related to the publication of the results, and etc.

Other database used in this study is the Medical Subject Headings (MeSH) thesaurus from U.S. National Library of Medicine (NLM). As the MeSH thesaurus revises and updates continually the medical terms, it is worth mentioning that the file 2017 MeSH Tree Hierarchy was downloaded from

NLM website on 10/23/2016. Based on MeSH thesaurus, we classify the diseases of the clinical trials in 30 classes and 363 subclasses similar to the first level and second level of the Anatomical Therapeutic Chemical (ATC) Classification System published by the World Health Organization (WHO).

The historic probability of drug approval was extracted from PharmaProjects, a database from Citeline. This organization claims to be the world's most comprehensive source of real-time R&D intelligence for the pharmaceutical industry. Pharmaprojects tracks drugs pipeline since 1980, covering all stages of drug development from discovery through clinical trials to drug approval.

Although the AACT database reflects the data from ClinicalTrials.gov until March 27, 2016, the sample was restricted to the clinical trials started during 2006 to 2011. The International Committee of Medical Journal Editors (ICMJE) established that clinical trials initiated after July 1, 2005, must be registered in the ClinicalTrials.gov website to be considered for publication. The Food and Drug Administration Amendments Act (FDAAA) requires that clinical trials from phase II to IV in U.S. site started after September 27, 2007, or with a completion date later than 12/25/2007 to be registered in the ClinicalTrials.gov. Following the CTTI (2016) cautionary advice that "studies that began before the ICMJE requirement in July 2005 are less likely to be registered, especially if their results are unpublished (e.g., negative studies)", and the substantial increasing in the registration of clinical trials in the ClinicalTrials.gov website during the 2002 to 2005, we decided to exclude the clinical trials started before 2006. For the same reason of avoiding potential bias in the sample, we also decided to exclude clinical trials started after 2012. Without this exclusion, we would underestimate the duration of clinical trials in the sample, because the short duration clinical trials would be overrepresented and the long duration clinical trials would be underrepresented in the sample.

Furthermore, only completed or terminated clinical trial that test chemical or biological compound entered in the sample, that is, ongoing study, surgery, devices, psychotherapy, dietary supplement, and etc, were excluded from the sample. We also excluded clinical trials, in which the primary purpose is educational, counseling, training, supportive care (maximize comfort instead of cure a disease), and health services research (processes, management, organization or financing of health care), because the focus of this study is in protocols designed to use chemical or biological compounds to treat, prevent, diagnose, or screen diseases.

In summary, our final sample is composed by 20,839 clinical trials started during 2006 to 2011 distributed in phases I, II, III, and IV, and conducted by 1,380 firms and 1,787 NPOs.

2.4.1 Variable Definitions

2.4.1.1 Dependent Variable

"Duration". The dependent variable "Duration" measured by months is equal the "Study Completion Date" minus the "Study Start Date". The "Study Start Date" is the date when the enrollment to the protocol starts. The "Study Completion Date" is the final date of data collection. This measure excludes the duration of FDA's drug review, allowing a direct comparison between NPOs and Firms concerned the duration of clinical trials. The results of this study remain the same if natural logarithm of "Duration" is used.

2.4.1.2 Independent Variables

"Firm Principal Sponsor". The main independent variable "Firm Principal Sponsor" is equal to 1, if the principal sponsor of the clinical trial is a firm, and 0 if the principal sponsor of the clinical trial is a NPO. Principal sponsor is defined in US Public Law 110-85, Title VIII, Section 801, as the primary organization responsible for the implementation and data analysis of the clinical trial.

"Firm Collaborator". This variable is equal to 1, if the clinical trial has at least one firm collaborator, and 0 otherwise.

"NPO Collaborator". This variable is equal to 1, if the clinical trial has at least one NPO collaborator, and 0 otherwise.

"Firm # of CTs". The variable "Firm Principal Sponsor = 1" was split in three dummies variables based on the number of clinical trials in a specific development phase. One dummy for firms below the median (< 50th perc.). A second dummy for firms between the second and third quartile (50th to 75th). A third dummy for firms above the third quartile (> 75th perc).

"NPO # of CTs". The variable "Firm Principal Sponsor = 0" was broken in three dummies variables based on the number of clinical trials in a specific development phase. One dummy for NPOs below the median (< 50th perc.). A second dummy for NPOs between the second and third quartile (50th to 75th). A third dummy for NPOs above the third quartile (> 75th perc).

"PS.Firm". The variable "Firm Principal Sponsor = 1" was split in four dummies based on cooperation with other organization. One dummy "PS.Firm (Single)" for clinical trials developed by a firm principal sponsor without any cooperation. A second dummy "PS.Firm with Firm" for clinical trials developed by a firm principal sponsor with external cooperation of other firm. A third dummy "PS.Firm with NPO" for clinical trials developed by a firm principal sponsor with external cooperation of a NPO. A fourth dummy "PS.Firm with Firm & NPO" for clinical trials developed by a firm principal sponsor with external cooperation of a firm and NPO.

"PS.NPO". The variable "Firm Principal Sponsor = 1" was split in four dummies based on cooperation with other organization. One dummy "PS.NPO (Single)" for clinical trials developed by a firm principal sponsor without any cooperation. A second dummy "PS.NPO with Firm" for clinical trials developed by a NPO principal sponsor with external cooperation of a firm. A third dummy "PS.NPO with NPO" for clinical trials developed by a NPO principal sponsor with external cooperation of other NPO. A fourth dummy "PS.NPO with Firm & NPO" for clinical trials developed by a firm principal sponsor with external cooperation of a firm and NPO.

2.4.1.3 Control Variables

"Data Monitoring Committee". This variable is equal to 1, if the clinical trial has data monitoring committee, and 0 otherwise. Data monitoring committee is an independent group of experts who reviews the ongoing data from clinical trials and advises the principal sponsor about the safety and efficacy of the intervention.

"Biological Compound". This variable is equal to 1, if the compound is biological, and 0 if the compound is chemical.

"# of CTs per Year". It is the number of clinical trials (all phases) per year at organization level.

"Ln(Enrollment)". It is the natural logarithm of the number of subjects in the trial.

"# of Countries". It is the number of different countries that hosted the clinical trial.

"Prob. of Success". It is the historic probability of drug approval. First, we calculate the probability of drug approval from 1989 to 2005 for 14 ATC level 1: ATC_A (Alimentary tract and metabolism), ATC_B (Blood and blood forming organs), ATC_C (Cardiovascular system), ATC_D (Dermatologicals), ATC_G (Genito-urinary system and sex hormones), ATC_H (Systemic hormonal preparations), ATC_J (Antiinfectives for systemic use), ATC_L (Antineoplastic and im-

munomodulating agents), ATC_M (Musculo-skeletal system), ATC_N (Nervous system), ATC_P (Antiparasitic products), ATC_R (Respiratory system), ATC_S (Sensory organs), and ATC_V (Various). Second, we compute the mean among the ATC1, in which the drug is classified.

"Phase I and Phase II". This variable is equal to 1, if phase I and phase II of the clinical trial are occurring in the same time under the same protocol, and 0 otherwise.

"Phase II and Phase III". This variable is equal to 1, if phase II and phase III of the clinical trial are occurring in the same time under the same protocol, and 0 otherwise.

"CT Goals". It is a set of dummy variables that captures the main goal of the clinical trial: 1) "Basic Science, 2) "Diagnostic", 3) "Screening", 4) "Prevention", and 5) "Treatment". The first is a protocol designed to investigate, for example the physiological or biomechanical mechanism of action of the compound in the human body. The second is a protocol designed to establish the presence of a disease. The third protocol aims to identify risk factors in large numbers of seemingly healthy people. The fourth protocol evaluates how a specific intervention prevents the development of a disease. The fifth is a protocol designed to assess how a specific intervention treat a disease.

"CT Design". It is composed by a set of variables related to the design of clinical trials: 1) "Minimum Age" - minimum age of participants measured in years; 2) "Maximum Age" - maximum age of participants measured in years; 3) "Male" - coded as 1, if only males are recruited, and 0 otherwise; 4) "Female" - coded as 1 if only females are recruited, and 0 otherwise; 5) "Male and Female" - coded as 1 if both males and females are recruited, 0 otherwise; 6) "Randomized" - coded as 1, if the participants are assigned to intervention groups at random, and 0 otherwise; 7) "# Intervention" - total number of interventions (biological, chemical, behavioral, device, dietary supplement, radiation, and etc); 8) "Open Label" - coded as 1, if no masking is used, and 0 otherwise; 9) "Single Blind" - coded as 1, if one party (the investigator or participant) is uninformed about the intervention assignment, and 0 otherwise; 10) "Double Blind" - coded as 1, if both parties are uninformed about the intervention assignment, and 0 otherwise; 11) "Single Group" - coded as 1, if it is single arm clinical trial, and 0 otherwise; 12) "Parallel" - coded as 1, if participants are assigned in parallel to one of two or more groups for the duration of the clinical trial, and 0 otherwise; 13) "Cross-over" - coded as 1, if participants switch between alternative interventions during the clinical trial, and 0 otherwise; 14) "Factorial" - coded as 1, if two or more interventions are evaluated in parallel against a control group, and 0 otherwise; 15) "Terminated" - coded as 1, if the clinical trial

finished prematurely the recruiting of participants, and 0 otherwise.

"CT Endpoint". It is a set of dummy variables that captures what outcome the protocol is designed to assess: 1) "Safety", 2) "Efficacy", 3) "Safety and Efficacy", 4) "Bio-equivalence", 5) "Bio-availability", 6) "Pharmacokinetics", 7) "Pharmacodynamics", 8) "Pharmacokinetics and Pharmacodynamics". The first endpoint examines how safe is the compound. The second endpoint quantifies the efficacy of the compound in healing the disease. The third endpoint is the combination of both first and second endpoints. The fourth endpoint checks if the generic compound is equivalent for all intents and purposes to the brand name compound. The fifth endpoint measures the rate and extent of the compound absorption. The sixth endpoint evaluates how the organism affects the compound over time in terms of absorption, distribution, biotransformation, and excretion. The seventh endpoint evaluates how the compound affects the organism over time in terms of intensity of therapeutic properties and adverse effects. The eighth endpoint is the combination of both seventh and sixth endpoints.

"Top 20 Countries". It is a set of dummy variables for 20 countries with the highest number of clinical trials in the sample. Following the countries in increasing order of number of clinical trials: Japan, Sweden, Taiwan, Switzerland, Poland, Brazil, Australia, Denmark, Israel, Belgium, Netherlands, South Korea, China, Spain, Italy, United Kingdom, France, Germany, Canada, and United States.

"Disease (30 classes)". It is a set of dummy variables for 30 classes of disease. The classes are: C01 - Bacterial Infections and Mycoses, C02 - Virus Diseases, C03 - Parasitic Diseases, C04 - Neoplasms, C05 - Musculoskeletal Diseases, C06 - Digestive System Diseases, C07 - Stomatognathic Diseases, C08 - Respiratory Tract Diseases, C09 - Otorhinolaryngologic Diseases, C10 - Nervous System Diseases, C11 - Eye Diseases, C12 - Male Urogenital Diseases, C13 - Female Urogenital Diseases and Pregnancy Complications, C14 - Cardiovascular Diseases, C15 - Hemic and Lymphatic Diseases, C16 - Congenital, Hereditary, and Neonatal Diseases and Abnormalities, C17 - Skin and Connective Tissue Diseases, C18 - Nutritional and Metabolic Diseases, C19 - Endocrine System Diseases, C20 - Immune System Diseases, C21 - Disorders of Environmental Origin, C22 - Animal Diseases, C23 - Pathological Conditions, Signs and Symptoms, C24 - Occupational Diseases, C25 - Chemically-Induced Disorders, C26 - Wounds and Injuries, F01 - Behavior and Behavior Mechanisms, F02 - Psychological Phenomena and Processes, F03 - Mental Disorders, F04 -

Behavioral Disciplines and Activities.

"Disease (363 subclasses)". It is a set of dummy variables for 363 subclasses of disease. Given the space restriction, some examples of subclasses will be provided, but all subclasses can be find in the MeSH thesaurus website. In the class C02 - Virus Diseases, one subclass is C02_440 - Hepatitis, Viral, Human. In the class C19 - Endocrine System Diseases, one subclass is C19_246 - Diabetes Mellitus. In the class C22 - Animal Diseases, one subclass is C22_812 - Salmonella Infections.

"Year". It is a set of dummy variables for each starting year of clinical trials from 2006 to 2011.

2.4.2 Summary Statistics

Table 2.1 presents the summary statistics of whole sample and the sample broken in principal sponsor NPOs and principal sponsor firm. Related to clinical trials conducted primary by NPOs, the average duration is 37 months, 26% receive external support by a firm, 15% receive external support by other NPOs, 58% have data monitoring committee, a typical NPO started on average 8 clinical trials per year during 2006 to 2011, on average 41 people enrolled per clinical trial, 6% of clinical trials are a combination of phase I and II, and 3% of clinical trials are a combination of phase II and III. Related to clinical trials conducted primary by firms, the average duration is 22 months, 10% receive external support by other firm, 4% receive external support by a NPO, 30% have data monitoring committee, a typical firm started on average 48 clinical trials per year during 2006 to 2011, on average 103 people enrolled per clinical trial, 4% of clinical trials are a combination of phase I and II, and 2% of clinical trials are a combination of phase II and III. Only the proportion of biological compound (11% to 13%) and the probability of drug approval (5.7% to 5.9%) are similar between NPOs and firms.

Figure 2.1 presents the proportion of clinical trials by organization and external cooperation. The most part of clinical trials has a firm as principal sponsor, except phase IV, in which NPOs as principal sponsor conduct more than 60% of clinical trials. It is also worth noting that the proportion of clinical trials conducted by firms as principal sponsor has been increasing during 2006 to 2011, except for phase IV. For all phases, about 20% of clinical trials has the external support by a firm or NPO, and this proportion shows a slightly decline over 2006 to 2011.

Table 2.2 presents the duration of clinical trials by phase. For each phase, firms complete the clinical trials early than the NPOs. Overall, the mean difference decreases from phase I to IV.

Table 2.1: Summary Statistics

	Whole Sample			Principal Sponsor NPO			Principal Sponsor Firm		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Duration (months)	20839	28.85	(20.73)	8910	37.25	(21.56)	11929	22.57	(17.64)
Firm Principal Sponsor	20839	0.57	(0.49)	8910	0.00	(0.00)	11929	1.00	(0.00)
Firm Collaborator	20839	0.17	(0.38)	8910	0.26	(0.44)	11929	0.10	(0.30)
NPO Collaborator	20839	0.15	(0.36)	8910	0.31	(0.46)	11929	0.04	(0.19)
Data Monitoring Committee	17221	0.43	(0.50)	7936	0.58	(0.49)	9285	0.30	(0.46)
Biological Compound	20839	0.12	(0.33)	8910	0.11	(0.31)	11929	0.13	(0.34)
# of CTs per Year	20839	31.57	(48.02)	8910	8.68	(13.70)	11929	48.66	(56.61)
Ln(Enrollment)	20784	4.25	(1.44)	8886	3.72	(1.28)	11898	4.64	(1.42)
Prob. of Success	20839	5.81	(2.09)	8910	5.68	(2.07)	11929	5.90	(2.10)
Phase I and Phase II	20839	0.06	(0.23)	8910	0.08	(0.27)	11929	0.04	(0.20)
Phase II and Phase III	20839	0.03	(0.16)	8910	0.04	(0.20)	11929	0.02	(0.13)

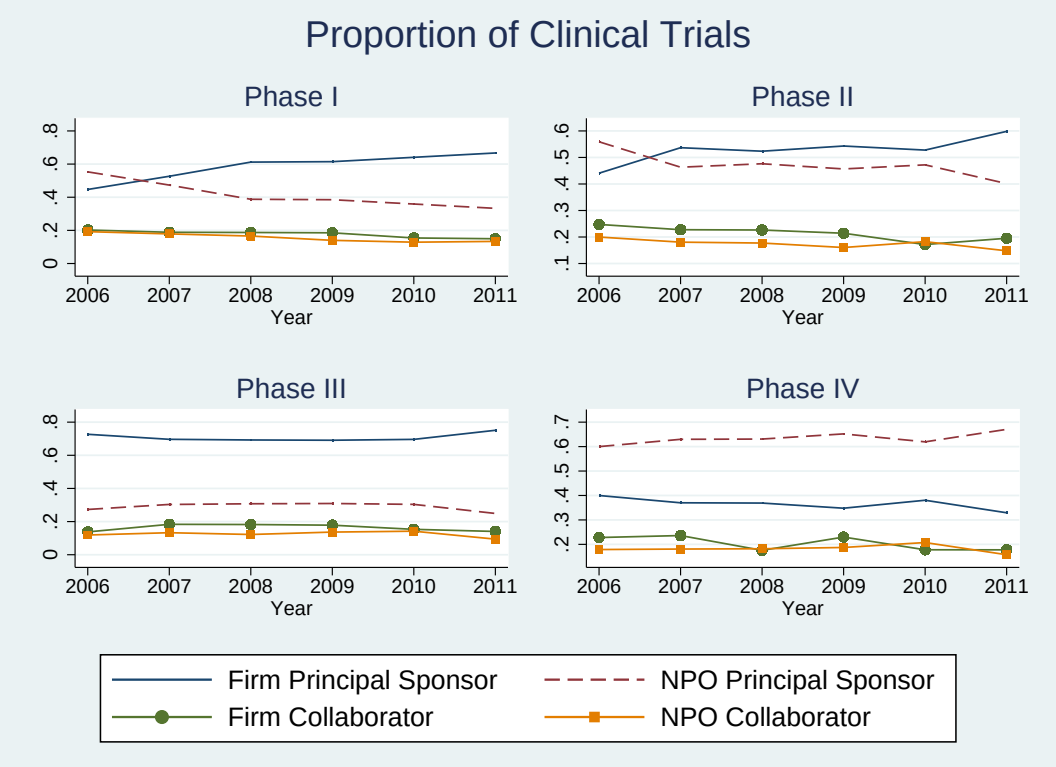


Figure 2.1: Proportion of Clinical Trials by Organization and External Cooperation

Phase I has the highest mean difference (20 months), but this difference has to be interpreted with caution. Firms are less likely to report the results of phase I, specially if the compound doesn't meet the safety standard. Compared to firms, NPOs are more likely to report the results of phase I, specially if they receive public funding. The mean difference of phase II is 19 months. A number relatively high if compared to the mean difference of combined clinical trials of phases I and II (12 months), and combined clinical trials of phases II and III (9 months). The mean difference of phase III is 13 months. In both phase II and III, the goal is to prove the effectiveness of the compound for a particular disease. The difference is that phase III demands a big trial (higher number of enrollments) and additional information about the risk-benefit of the compound. One explanation for lower mean difference of phase III in comparison to phase I and II is that optimizing the protocol of phase III involves higher risk in terms of receiving drug approval than optimizing the duration of tasks and procedures of phase I and II. Phase IV has the lowest mean difference (8 months). The goal of phase IV is to obtain extra information about compound's side effects, benefits, and optimal use. Almost always, clinical trials, that aims to identify uncommon adverse

Table 2.2: Duration of Clinical Trials by Phase (Measured in Months)

	Principal Sponsor NPO			Principal Sponsor Firm		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Phase I	1258	37.32	(22.09)	2352	17.49	(17.73)
Phase I and Phase II	700	40.14	(23.01)	518	28.41	(19.62)
Phase II	3163	41.34	(21.71)	3980	22.69	(17.15)
Phase II and Phase III	369	35.95	(20.19)	203	26.16	(16.51)
Phase III	1188	37.50	(22.12)	3583	24.91	(17.57)
Phase IV	2232	30.60	(18.66)	1293	22.04	(16.23)
Total	8910	37.25	(21.56)	11929	22.57	(17.64)

effects in special populations, are relegate to phase IV [Juni et al. (2001), Friedman et al. (2015)]. It is worth noting that firms have less incentives to optimize the duration and carry out clinical trials after the FDA drug approval. Phase IV represents only 10% of clinical trials conducted by firms, but 25% of clinical trials conducted by NPOs. Firms appear to focus on phases II and III; whereas NPOs appear to focus on phase II and IV.

Table 2.3 presents the duration of clinical trials by organization and external cooperation. The categories "PS.Firm (Single)", "PS.Firm with Firm", "PS.Firm with NPO", "PS.Firm with Firm & NPO", "PS.NPO (Single)", "PS.NPO with Firm", "PS.NPO with NPO", and "PS.NPO with Firm & NPO" are mutually exclusive. Overall, for each phase: 1) firm principal sponsor completes clinical trials early than NPO principal sponsor; 2) firm principal sponsor without any external support completes clinical trials early than other categories; 3) any kind of external cooperation (firm, NPO, and both) appears to delay the clinical trials in comparison with the situation of single development, or there is not statistically significance difference between cooperation and not cooperation.

Table 2.4 presents the correlation matrix of main variables used in the regressions of Tables 2.5, 2.6, 2.7, and 2.8. Overall, the sign of the variables is aligned to the hypothesis of section 2.3 and results of section 2.6.

2.5 Empirical Strategy

The goal of the empirical strategy is to identify the causal impact of the variable "Firm Principal Sponsor" on the duration of clinical trials. The data was not originated by a randomized experiment. However, given the period analyzed, legal requirement by regulatory authority, norms about the

Table 2.3: Duration of Clinical Trials by External Cooperation (Measured in Months)

	Phase I		Phase II		Phase III		Phase IV	
	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean
PS.Firm (Single)	2477	19.13	4051	23.08	3230	24.52	1137	21.87
PS.Firm with Firm	247	20.70	430	25.55	459	26.57	106	22.21
PS.Firm with NPO	135	23.21	192	27.84	76	31.97	40	24.90
PS.Firm with Firm & NPO	11	19.18	28	19.18	21	35.43	10	28.40
PS.NPO (Single)	964	36.21	1907	39.13	756	34.64	1157	28.91
PS.NPO with Firm	399	43.34	977	42.43	233	35.65	479	29.70
PS.NPO with NPO	486	38.12	1066	41.50	483	40.60	514	34.28
PS.NPO with Firm & NPO	109	39.68	282	41.90	85	43.66	82	36.67
Total	4828	27.11	8933	31.62	5343	28.52	3525	27.46

Table 2.4: Correlation Matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Duration (months)	(1)	1.00										
Firm Principal Sponsor	(2)	-0.35**	1.00									
Firm Collaborator	(3)	0.11**	-0.21**	1.00								
NPO Collaborator	(4)	0.18**	-0.37**	0.00	1.00							
Data Monitoring Committee	(5)	0.25**	-0.28**	0.05**	0.21**	1.00						
Biological Compound	(6)	0.01	0.04**	-0.05**	0.04**	0.10**	1.00					
# of CTs per Year	(7)	-0.15**	0.41**	-0.13**	-0.21**	-0.17**	0.08**	1.00				
Ln(Enrollment)	(8)	0.00	0.32**	-0.09**	-0.07**	0.01	0.05**	0.19**	1.00			
Prob. of Success	(9)	-0.12**	0.05**	-0.02**	-0.04**	-0.06**	-0.03**	-0.01	0.04**	1.00		
Phase I and Phase II	(10)	0.08**	-0.07**	0.03**	0.05**	0.06**	0.07**	-0.09**	-0.15**	-0.03**	1.00	
Phase II and Phase III	(11)	0.03**	-0.07**	0.01*	0.05**	0.05**	-0.01	-0.06**	0.03**	0.01	-0.04**	1.00

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

publication of results, and economic importance, the sample of phase III is likely to be representative of the population. Other phases are more likely to suffer from sampling bias. Specially for phase I, some studies with undesirable results conducted by small organizations might not be registered in the ClinicalTrials.gov website. Therefore, the results of this paper are more reliable for phase III, and causal impact has to be interpreted as holding all other factors constant.

We propose the following model:

$$y = \alpha D + x^{'}\beta + u \tag{2.1}$$

where y is the duration of clinical trial, D is the dummy variable "Firm Principal Sponsor", x' is the vector of control variables, and u is error term. In order to estimate the coefficients α and β, we assume that y_0 (outcome for the non-treated group) and y_1(outcome for the treated group) are $\perp$ (independent) of D conditional to the control variables x:

$$y_0, y_1 \perp D|x \tag{2.2}$$

Equation 2.2 implies that although the data is not originated by an experimental setting, the problem of confounding factors or selection bias can be eliminated once x is included in the regression. Then, equation 2.1 can be estimated by ordinary least squares (OLS), clustering the robust standard errors at organization level (firms and NPOs).

To the best of our knowledge, this study has the most comprehensive list of control variables. The ideal clinical trial has two characteristics: randomized and double-blind. In order to control for the quality of clinical trials [Juni et al. (2001), Friedman et al. (2015)], we have the variables "Randomized", "Open Label", "Single Blind", "Double Blind", and "Data Monitoring Committee". Medical literature [Fleming and DeMets (1996), Johnson et al. (2003), Wagner et al. (2007)] points outs that duration and rate of success of clinical trials are very sensitive to the clinical endpoint or surrogates. Unfortunately, we don't control for a specific endpoint as survival rate, or surrogates as tumor shrinkage, cholesterol levels, and blood pressure, but we control for general categories of endpoints with the variables: "Safety", "Efficacy", "Bio-equivalence", "Bio-availability", "Pharmacokinetics", "Pharmacodynamics" and interactions of previous variables. Budish et al. (2015) also point out that firms prefer to develop late-stage cancer drugs than cancer prevention drugs, because

the first has higher expected payoff and can be brought to the market faster; whereas the second involves longer clinical trials and higher discount factor. We control for these factors with the variables: "Prevention" and "Treatment". Isakov et al. (2016) suggest that for terminal diseases, FDA is too conservative with drug approval requirements; but for less deadly illnesses, FDA is too risk-tolerant. As clinical trials conducted in different countries are subject to different regulations and standards, we control for this variation with dummies for "Top 20 Countries" in terms of quantity of clinical trials. Furthermore, we control for disease variation ("30 classes" and "363 subclasses"), and potential shocks over the years ("Year").

Therefore, we estimate the following equation:

$$Duration_{cody} = \alpha_1 Firm\_Principal\_Sponsor_{cody} + \alpha_2 Firm\_Collaborator_{cody}$$
$$+\alpha_3 NPO\_Collaborator_{cody} + x'_{cody}\beta + \kappa_c + \delta_d + \gamma_y + u_{cody} \tag{2.3}$$

The dependent variable, $Duration_{cody}$, is the duration of clinical trial c, conducted by organization o, targeting disease d, starting in the year y. The main independent variable is the dummy variable "Firm Principal Sponsor". The secondary independent variables are the dummy variables for "Firm Collaborator" and "NPO Collaborator". Variables for clinical trial goals, design, and endpoint are represented by κ_c, dummies for specific diseases (30 classes or 363 subclasses) are represented by δ_d, and dummies for starting year of clinical trials are represented by γ_y.

For robustness check, we estimate α_1 using the Nearest-Neighbor Matching estimator (NNME) with robust standard errors derived by Abadie and Imbens (2006, 2012). Nearest-Neighbor Matching is a nonparametric estimator that imposes minimal functional form and exclusion restriction for identification, but requires bigger sample size than OLS. According with Cameron and Trivedi (2005), the main advantage of the matching estimator in comparison to OLS is to relax the assumption that the variable of interest D enters the conditional mean function linearly.

In addition to assumption 2 that is critical to identify the population-average treatment effect on the treated (ATET), we assume the overlap assumption:

$$Pr[D = 1|x] < 1 \tag{2.4}$$

Equation 2.4 guarantees that for each clinical trial conducted by a firm, there is another clinical

trial conducted by NPO with similar x. As only y_{0i} or y_{1i} is observed, but not both in the same time for an organization i, the matching estimator estimates the missing potential outcome based on the outcomes of the "nearest neighbors". In other words, the NNME imputes the missing values by matching firms and NPOs that are most similar based on the vector of control variables x. To calculate the distance between x_i and x_j, Mahalanobis distance, a weight that uses the inverse of the covariates' variance–covariance matrix, is employed [Bloom et al. (2013)]. The results remain the same if inverse variance or Euclidean metric is used.

Related to the number of matches per observation, there is a trade off between bias and variance [Cameron and Trivedi (2005)]. Although higher number of matches per observation decreases the variance, we decided for only one match per observation to minimize the probability of inferior match. Instead of using propensity score, we also decided to match on the regressors, because only few variables in x are continuous. In the presence of continuous variables, the NNME estimator is biased even in large samples [Abadie and Imbens (2006, 2012)]. Therefore, we corrected for bias based on Abadie and Imbens (2006, 2012), using a specific linear function of the continuous variables: "Ln(Enrollment)", "Minimum Age", "Maximum Age", and "Prob. of Success".

For additional robustness check, we also estimate the impact of "Firm Principal Sponsor" on clinical trial hazard rate (completion of clinical trial), using the Cox proportional hazard model with robust standard errors clustered by organizations. Cox model is a semiparametric method [Cox (1972), Wooldridge (2010)]. The proportional hazard is composed by a baseline hazard common to all clinical trials $\lambda_0(t)$ that is function of time t, and individual hazard functions $\mu(x\beta)$ that differ proportionately based on regressors x and vector of parameters β:

$$\lambda(t; x\beta) = \mu(x\beta)\lambda_0(t) \tag{2.5}$$

Cox model is appropriate, when the regressors are time-invariant, and someone is interested in how the regressors shift the hazard function [Wooldridge (2010)]. The advantage of Cox model is that the hazard ratios of the covariates can be estimated very generally without strict assumptions about $\lambda_0(t)$. Cox (1972) derived the partial maximum likelihood estimator for β without the need to estimate $\lambda_0(t)$.

Therefore, we investigate the impact of the variable "Firm Principal Sponsor" on the duration

of clinical trials, using the OLS as baseline approach and two other methods for robustness check.

2.6 Result

Table 2.5 presents the OLS estimation of duration of phase I. Overall, the results support H1, as the coefficient of "Firm Principal Sponsor" is negative and statistically significant at 1% level in the models 1, 2 and 5. Controlling for several factors, the baseline model 2 estimates that clinical trials conducted by firms finish 13 months early than clinical trials conducted by NPO. This result persists in model 3, as the coefficients of "Firm # of CTs (< 50th perc.)", "Firm # of CTs (50th to 75th)", and "Firm # of CTs (> 75th perc.)" are negative and statistically significant at 1% level. As the three coefficients have similar magnitude, we infer that previous results were not driven by firm size or experience in clinical trials. By the other hand, medium (50th to 75th) and bigger (> 75th perc.) NPOs with more experience take about additional 4 months to finish a clinical trial in comparison to smaller NPO with less experience (< 50th perc.). It looks that bureaucracy related to size affects more the NPOs than firms. Concerned H2, there is empirical evidence that an external NPO collaboration delays the clinical trials, as indicated by the positive and statistically significant coefficient of "NPO Collaborator" at 1% level in the models 2, 3 and 5. According with baseline model 2, external NPO collaboration delays the clinical trial in 2 months. However, external firm collaboration has no effect on the duration of clinical trials as indicated by the non-statistically significant coefficient of "Firm Collaborator" in the models 2, 3, and 5. Model 4 does not support the relative efficiency of external firm collaboration in comparison to external NPO collaboration. Wald tests indicate that: I) "PS.Firm with Firm" is not statistically different than "PS.Firm with NPO" at 10% level; and II) "PS.NPO with Firm" is not statistically different than "PS.NPO with NPO" at 10% level. Furthermore, "PS.Firm with Firm & NPO" and "PS.NPO with Firm & NPO" have the same impact on the duration of the clinical trials as respectively "PS.Firm (Single)" and "PS.NPO (Single)". For robustness checks, model 5 confirms the results of previous model, but with slightly lower magnitude of the coefficients due to additional control variables ("Data Monitoring Committee" and "363 diseases subclasses") that remove part of variation.

Table 2.6 presents the OLS estimation of duration of phase II. Overall, the results support H1, as the coefficient of "Firm Principal Sponsor" is negative and statistically significant at 1% level in the

Table 2.5: OLS Estimation of Duration of Phase I

	(1)	(2)	(3)	(4)	(5)
Firm Principal Sponsor	-16.59***	-13.25***			-11.70***
	(1.17)	(1.21)			(1.12)
Firm Collaborator		0.73	0.66		1.13
		(1.09)	(1.09)		(1.08)
NPO Collaborator		2.21*	2.36**		2.16**
		(1.19)	(1.07)		(1.09)
Firm # of CTs (< 50th perc.)			-11.42***		
			(1.12)		
Firm # of CTs (50th to 75th)			-9.30***		
			(1.99)		
Firm # of CTs (> 75th perc.)			-10.43***		
			(2.43)		
NPO # of CTs (50th to 75th)			4.97**		
			(2.06)		
NPO # of CTs (> 75th perc.)			4.18**		
			(2.02)		
PS.Firm (Single)				-12.61***	
				(1.60)	
PS.Firm with Firm				-12.78***	
				(1.91)	
PS.Firm with NPO				-10.59***	
				(2.15)	
PS.Firm with Firm & NPO				-12.81***	
				(3.66)	
PS.NPO with Firm				2.74	
				(2.16)	
PS.NPO with NPO				3.24*	
				(1.85)	
PS.NPO with Firm & NPO				1.88	
				(3.05)	
Data Monitoring Committee					2.42***
					(0.85)
Biological Compound		4.26***	4.12***	4.30***	4.55***
		(1.11)	(1.11)	(1.11)	(1.30)
# of CTs per Year		-0.01**	-0.02	-0.01**	-0.00
		(0.01)	(0.02)	(0.01)	(0.01)
Ln(Enrollment)		3.77***	3.84***	3.78***	3.79***
		(0.45)	(0.46)	(0.45)	(0.41)
Prob. of Success		-0.51	-0.46	-0.46	-1.15***
		(0.38)	(0.36)	(0.39)	(0.39)
Phase I and Phase II		2.41***	2.60***	2.43***	3.10***
		(0.86)	(0.85)	(0.86)	(0.99)
CT Goals	No	Yes	Yes	Yes	Yes
CT Design	No	Yes	Yes	Yes	Yes
CT Endpoint	No	Yes	Yes	Yes	Yes
Top 20 Countries (# of CTs)	Yes	Yes	Yes	Yes	Yes
Disease (30 classes)	Yes	Yes	Yes	Yes	No
Disease (363 subclasses)	No	No	No	No	Yes
Year	Yes	Yes	Yes	Yes	Yes
R^2	0.40	0.47	0.48	0.47	0.51
Observations	4828	3031	3031	3031	2507

Note: Robust standard errors are in parentheses (clustered by organizations).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

models 1, 2 and 5. Controlling for several factors, the baseline model 2 estimates that clinical trials conducted by firms finish 14 months early than clinical trials conducted by NPO. This result persists in model 3, as the coefficients of "Firm # of CTs (< 50th perc.)", "Firm # of CTs (50th to 75th)", and "Firm # of CTs (> 75th perc.)" are negative and statistically significant at 1% level. As the three coefficients have similar magnitude, we infer that previous results were not driven by firm size or experience in clinical trials. By the other hand, medium (50th to 75th) and bigger (> 75th perc.) NPOs with more experience take respectively additional 3 and 4 months to finish a clinical trial in comparison to smaller NPO with less experience (< 50th perc.). It looks that bureaucracy related to size affects more the NPOs than firms. Concerned H2, there is empirical evidence that an external NPO collaboration delays the clinical trials, as indicated by the positive and statistically significant coefficient of "NPO Collaborator" at 1% level in the models 2, 3 and 5. According with baseline model 2, external NPO collaboration delays the clinical trial in 5 months. However, external firm collaboration has no effect on the duration of clinical trials as indicated by the non-statistically significant coefficient of "Firm Collaborator" in the models 2, 3, and 5. Model 4 supports the relative efficiency of external firm collaboration in comparison to external NPO collaboration. Wald tests indicate that: I) "PS.Firm with Firm" is statistically different than "PS.Firm with NPO" at 1% level; and II) "PS.NPO with Firm" is statistically different than "PS.NPO with NPO" at 1% level. Furthermore, combined cooperation (firm and NPO) speeds up the clinical trials in comparison to single NPO cooperation. For robustness checks, model 5 confirms the results of previous model, but with slightly lower magnitude of the coefficients due to additional control variables ("Data Monitoring Committee" and "363 diseases subclasses") that remove part of variation.

Table 2.7 presents the OLS estimation of duration of phase III. Overall, the results support H1, as the coefficient of "Firm Principal Sponsor" is negative and statistically significant at 1% level in the models 1, 2 and 5. Controlling for several factors, the baseline model 2 estimates that clinical trials conducted by firms finish 16 months early than clinical trials conducted by NPO. This result persists in model 3, as the coefficients of "Firm # of CTs (< 50th perc.)", "Firm # of CTs (50th to 75th)", and "Firm # of CTs (> 75th perc.)" are negative and statistically significant at 1% level. As the three coefficients have similar magnitude, we infer that previous results were not driven by firm size or experience in clinical trials. By the other hand, bigger NPOs with more experience (> 75th perc.) takes additional 5 months to finish a clinical trial in comparison to smaller NPO with less experience

Table 2.6: OLS Estimation of Duration of Phase II

	(1)	(2)	(3)	(4)	(5)
Firm Principal Sponsor	-16.44***	-14.57***			-13.85***
	(0.71)	(1.03)			(0.83)
Firm Collaborator		1.00	0.93		0.94
		(0.80)	(0.77)		(0.78)
NPO Collaborator		5.35***	5.42***		4.50***
		(1.00)	(0.95)		(0.86)
Firm # of CTs (< 50th perc.)			-12.85***		
			(0.98)		
Firm # of CTs (50th to 75th)			-12.20***		
			(1.50)		
Firm # of CTs (> 75th perc.)			-11.43***		
			(2.09)		
NPO # of CTs (50th to 75th)			3.29***		
			(1.13)		
NPO # of CTs (> 75th perc.)			4.79***		
			(1.75)		
PS.Firm (Single)				-14.71***	
				(1.36)	
PS.Firm with Firm				-12.51***	
				(1.61)	
PS.Firm with NPO				-7.25***	
				(1.86)	
PS.Firm with Firm & NPO				-14.08***	
				(2.44)	
PS.NPO with Firm				1.61	
				(1.54)	
PS.NPO with NPO				5.92***	
				(1.49)	
PS.NPO with Firm & NPO				3.38*	
				(1.83)	
Data Monitoring Committee					3.38***
					(0.57)
Biological Compound		3.16***	2.92***	3.17***	4.61***
		(1.05)	(1.08)	(1.04)	(0.96)
# of CTs per Year		0.00	-0.01	0.00	0.01**
		(0.01)	(0.01)	(0.01)	(0.00)
Ln(Enrollment)		1.91***	1.95***	1.91***	2.25***
		(0.27)	(0.28)	(0.28)	(0.30)
Prob. of Success		-0.37	-0.30	-0.35	-0.79***
		(0.34)	(0.33)	(0.34)	(0.30)
Phase I and Phase II		1.88**	1.76**	1.82**	1.91**
		(0.82)	(0.83)	(0.81)	(0.87)
Phase II and Phase III		1.79*	2.03**	1.71*	2.40**
		(0.95)	(0.92)	(0.95)	(1.01)
CT Goals	No	Yes	Yes	Yes	Yes
CT Design	No	Yes	Yes	Yes	Yes
CT Endpoint	No	Yes	Yes	Yes	Yes
Top 20 Countries (# of CTs)	Yes	Yes	Yes	Yes	Yes
Disease (30 classes)	Yes	Yes	Yes	Yes	No
Disease (363 subclasses)	No	No	No	No	Yes
Year	Yes	Yes	Yes	Yes	Yes
R^2	0.30	0.33	0.34	0.33	0.40
Observations	8933	6137	6137	6137	5236

Note: Robust standard errors are in parentheses (clustered by organizations).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(< 50th perc.). It looks that bureaucracy related to size affects more the NPOs than firms. Concerned H2, there is empirical evidence that an external NPO collaboration delays the clinical trials, as indicated by the positive and statistically significant coefficient of "NPO Collaborator" at 1% level in the models 2, 3 and 5. According with baseline model 2, external NPO collaboration delays the clinical trial in 6 months. However, external firm collaboration has no effect on the duration of clinical trials as indicated by the non-statistically significant coefficient of "Firm Collaborator" in the models 2, 3, and 5. Model 4 supports the relative efficiency of external firm collaboration in comparison to external NPO collaboration. Wald tests indicate that: I) "PS.Firm with Firm" is statistically different than "PS.Firm with NPO" at 1% level; and II) "PS.NPO with Firm" is statistically different than "PS.NPO with NPO" at 5% level. Furthermore, combined cooperation (firm and NPO) delays the clinical trials in comparison to single development by firm or NPO. For robustness checks, model 5 confirms the results of previous model, but with slightly lower magnitude of the coefficients due to additional control variables ("Data Monitoring Committee" and "363 diseases subclasses") that remove part of variation.

Table 2.8 presents the OLS estimation of duration of phase IV. Overall, the results support H1, as the coefficient of "Firm Principal Sponsor" is negative and statistically significant at 1% level in the models 1, 2 and 5. Controlling for several factors, the baseline model 2 estimates that clinical trials conducted by firms finish 11 months early than clinical trials conducted by NPO. This result persists in model 3, as the coefficients of "Firm # of CTs (< 50th perc.)", "Firm # of CTs (50th to 75th)", and "Firm # of CTs (> 75th perc.)" are negative and statistically significant at 1% level. However, bigger firms and NPOs with more experience (> 75th perc.) take longer to finish clinical trials in comparison to smaller (< 50th perc.) and medium (50th to 75th) size firms and NPO with less experience. Concerned H2, there is empirical evidence that an external NPO collaboration delays the clinical trials, as indicated by the positive and statistically significant coefficient of "NPO Collaborator" at 1% level in the models 2, 3 and 5. According with baseline model 2, external NPO collaboration delays the clinical trial in 4 months. However, external firm collaboration has no effect on the duration of clinical trials as indicated by the non-statistically significant coefficient of "Firm Collaborator" in the models 2, 3, and 5. Model 4 supports the relative efficiency of external firm collaboration in comparison to external NPO collaboration. Wald tests indicate that: I) "PS.Firm with Firm" is statistically different than "PS.Firm with NPO" at 5%

Table 2.7: OLS Estimation of Duration of Phase III

	(1)	(2)	(3)	(4)	(5)
Firm Principal Sponsor	-16.86***	-16.43***			-14.23***
	(1.00)	(1.14)			(1.09)
Firm Collaborator		0.00	-0.09		0.75
		(0.90)	(0.89)		(0.86)
NPO Collaborator		6.05***	6.03***		5.16***
		(1.10)	(1.08)		(1.16)
Firm # of CTs (< 50th perc.)			-15.80***		
			(1.19)		
Firm # of CTs (50th to 75th)			-15.93***		
			(2.24)		
Firm # of CTs (> 75th perc.)			-16.01***		
			(3.32)		
NPO # of CTs (50th to 75th)			-1.17		
			(1.35)		
NPO # of CTs (> 75th perc.)			5.03***		
			(1.89)		
PS.Firm (Single)				-16.16***	
				(1.38)	
PS.Firm with Firm				-17.12***	
				(1.47)	
PS.Firm with NPO				-8.44***	
				(2.65)	
PS.Firm with Firm & NPO				-10.10**	
				(4.63)	
PS.NPO with Firm				1.48	
				(1.85)	
PS.NPO with NPO				5.79***	
				(1.42)	
PS.NPO with Firm & NPO				7.80***	
				(2.82)	
Data Monitoring Committee					4.88***
					(0.75)
Biological Compound		3.76***	3.76***	3.75***	4.20***
		(1.31)	(1.26)	(1.31)	(1.32)
# of CTs per Year		-0.03***	-0.03	-0.03***	-0.01
		(0.01)	(0.02)	(0.01)	(0.01)
Ln(Enrollment)		1.74***	1.80***	1.78***	1.89***
		(0.30)	(0.29)	(0.30)	(0.31)
Prob. of Success		0.65	0.65	0.66	0.47
		(0.43)	(0.43)	(0.43)	(0.37)
Phase II and Phase III		0.50	0.49	0.55	0.36
		(1.01)	(0.99)	(1.01)	(1.10)
CT Goals	No	Yes	Yes	Yes	Yes
CT Design	No	Yes	Yes	Yes	Yes
CT Endpoint	No	Yes	Yes	Yes	Yes
Top 20 Countries (# of CTs)	Yes	Yes	Yes	Yes	Yes
Disease (30 classes)	Yes	Yes	Yes	Yes	No
Disease (363 subclasses)	No	No	No	No	Yes
Year	Yes	Yes	Yes	Yes	Yes
R^2	0.27	0.33	0.33	0.33	0.39
Observations	5343	4570	4570	4570	3712

Note: Robust standard errors are in parentheses (clustered by organizations).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

level; and II) "PS.NPO with Firm" is statistically different than "PS.NPO with NPO" at 1% level. Furthermore, combined cooperation (firm and NPO) delays the clinical trials in comparison to single development by firm or NPO. For robustness checks, model 5 confirms the results of previous model, but with slightly lower magnitude of the coefficients due to additional control variables ("Data Monitoring Committee" and "363 diseases subclasses") that remove part of variation.

Table 2.9 presents the firm principal sponsor impact on the duration of clinical trial for each phase from I to IV based on nearest-neighbor matching estimator. All estimations have the same specification of model 2 (Tables 5, 6, 7, and 8). Robust standard errors were calculated based on Abadie and Imbens (2006, 2012). We also corrected for large-sample bias according with Abadie and Imbens (2006, 2012), using the variables: "Ln(Enrollment)", "Minimum Age", "Maximum Age", and "Prob. of Success". Overall the results of matching estimator are similar to the results of OLS estimator of Tables 5, 6, 7, and 8. All ATET (Mahalanobis Metric) are statistically significant at 1% level. The matching estimator predicts respectively for phases I, II, III, and IV, that firms finish clinical trial about 18.8, 16.3, 14.3, and 10.6 months early than NPO. In comparison to the OLS baseline model 2 of Tables 2.5, 2.6, 2.7, and 2.8 (respectively 13.2, 14.5, 16.4, and 11.4 months), the matching estimator estimates a higher impact of "Firm Principal Sponsor" for phase I and II, but lower impact for phase III and IV. The higher discrepancy between both estimators occurs in phase I, in which is more likely to suffer of sample bias as described in previous section.

Table 2.10 presents the Cox proportional hazard ratios for duration of clinical trials. All estimations have the same specification of model 2 (Tables 2.5, 2.6, 2.7, and 2.8). Robust standard errors were clustered by organizations. The results are qualitatively similar to OLS and matching estimators. The variable "Firm Principal Sponsor" is statistically significant at 1% level, and associated with a more than 100% increase in the hazard rate (completion of clinical trial) for all phases. In other words, for any point in time, firms are at least twice more likely to finish the clinical trials than NPOs for all phases. The variable "Firm Collaborator" is not statistically significant, but variable "NPO Collaborator" is statistically significant at 5% for phase I, and 1% for phases II, III, and IV. The variable "NPO Collaborator" is associated with about 20% decrease in the hazard rate, supporting the H2 that external support of NPO delays the completion of the clinical trials.

Table 2.8: OLS Estimation of Duration of Phase IV

	(1)	(2)	(3)	(4)	(5)
Firm Principal Sponsor	-10.09***	-11.39***			-11.20***
	(0.91)	(1.30)			(1.26)
Firm Collaborator		-0.62	-0.51		-0.65
		(0.88)	(0.89)		(0.92)
NPO Collaborator		4.63***	4.57***		3.48***
		(1.06)	(1.06)		(1.11)
Firm # of CTs (< 50th perc.)			-10.44***		
			(1.31)		
Firm # of CTs (50th to 75th)			-9.48***		
			(3.47)		
Firm # of CTs (> 75th perc.)			-6.55**		
			(3.00)		
NPO # of CTs (50th to 75th)			1.24		
			(1.22)		
NPO # of CTs (> 75th perc.)			2.29*		
			(1.29)		
PS.Firm (Single)				-11.62***	
				(1.40)	
PS.Firm with Firm				-12.59***	
				(1.71)	
PS.Firm with NPO				-4.16	
				(4.13)	
PS.Firm with Firm & NPO				-3.90	
				(3.91)	
PS.NPO with Firm				-0.58	
				(1.18)	
PS.NPO with NPO				4.40***	
				(1.24)	
PS.NPO with Firm & NPO				3.38*	
				(1.98)	
Data Monitoring Committee					2.40**
					(0.94)
Biological Compound		-1.84	-1.77	-1.81	1.28
		(1.65)	(1.67)	(1.66)	(1.87)
# of CTs per Year		-0.02*	-0.04*	-0.02*	-0.03*
		(0.01)	(0.02)	(0.01)	(0.02)
Ln(Enrollment)		2.23***	2.25***	2.24***	2.83***
		(0.41)	(0.41)	(0.41)	(0.41)
Prob. of Success		-0.33	-0.35	-0.35	-0.11
		(0.44)	(0.44)	(0.44)	(0.43)
CT Goals	No	Yes	Yes	Yes	Yes
CT Design	No	Yes	Yes	Yes	Yes
CT Endpoint	No	Yes	Yes	Yes	Yes
Top 20 Countries (# of CTs)	Yes	Yes	Yes	Yes	Yes
Disease (30 classes)	Yes	Yes	Yes	Yes	No
Disease (363 subclasses)	No	No	No	No	Yes
Year	Yes	Yes	Yes	Yes	Yes
R^2	0.15	0.21	0.22	0.21	0.30
Observations	3525	2669	2669	2669	2288

Note: Robust standard errors are in parentheses (clustered by organizations).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Firm Principal Sponsor Impact (Nearest-Neighbor Matching Estimates of ATET)

	(1) Phase I	(2) Phase II	(3) Phase III	(4) Phase IV
ATET (Mahalanobis Metric)	-18.79***	-16.32***	-14.31***	-10.66***
	(1.15)	(0.93)	(1.41)	(0.99)
Observations	3031	6137	4570	3411

Note 1: Robust standard errors derived by Abadie and Imbens (2006 and 2012) are in parentheses.

Note 2: Each phase has the same covariates of model 2 (Tables 2.5, 2.6, 2.7, and 2.8).

Note 3: All estimations were corrected for large-sample bias according with Abadie and Imbens (2006, 2012), using the variables: Ln(Enrollment), Minimum Age, Maximum Age, and Prob. of Success.

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 2.10: Cox Proportional Hazard Ratios for Duration of Clinical Trials

	(1) Phase I	(2) Phase II	(3) Phase III	(4) Phase IV
Firm Principal Sponsor	2.44***	2.30***	2.53***	2.02***
	(0.17)	(0.13)	(0.16)	(0.17)
Firm Collaborator	1.02	0.96	1.01	1.08
	(0.06)	(0.04)	(0.05)	(0.05)
NPO Collaborator	0.87**	0.75***	0.78***	0.78***
	(0.06)	(0.04)	(0.04)	(0.04)
Biological Compound	0.77***	0.79***	0.84**	0.99
	(0.05)	(0.06)	(0.06)	(0.09)
# of CTs per Year	1.00***	1.00	1.00***	1.00
	(0.00)	(0.00)	(0.00)	(0.00)
Ln(Enrollment)	0.78***	0.91***	0.92***	0.89***
	(0.02)	(0.02)	(0.02)	(0.02)
Prob. of Success	1.03	1.03*	0.98	1.02
	(0.03)	(0.02)	(0.03)	(0.03)
Phase I and Phase II	0.83***	0.89**		
	(0.04)	(0.04)		
Phase II and Phase III		0.86***	0.91	
		(0.05)	(0.05)	
CT Goals	Yes	Yes	Yes	Yes
CT Design	Yes	Yes	Yes	Yes
CT Endpoint	Yes	Yes	Yes	Yes
Top 20 Countries (# of CTs)	Yes	Yes	Yes	Yes
Disease (30 classes)	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Log Likelihood	-20264.84	-46421.46	-33267.29	-18111.08
Observations	3017	6132	4563	2662

Note: Robust standard errors are in parentheses (clustered by organizations).

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

2.7 Conclusion

This study provided empirical evidence that firms are more efficient in conducting clinical trials than NPOs. Specifically, based on our OLS baseline model, we estimated that firms finish clinical trials 13, 14, 16, and 11 months earlier than NPOs respectively for phases I, II, III, and IV. Other evidence that firms manage more efficiently clinical trials is that NPOs external support delays clinical trials in 2, 5, 6, and 4 months respectively for phases I, II, III, and IV, but firm external support doesn't appear to affect the duration of clinical trials for any phase. Therefore, we provide empirical support to Aghion et al. (2008) theoretical prediction that firms have comparative advantage in the late-stage research; whereas NPOs have comparative advantage in the early-stage research.

Although we run several specifications, control for several factors, and check our results with matching estimators, our study has some limitations that can be addressed in future research. First, this study represents a static picture of relative efficiency as the sample is restricted to the clinical trials with starting year from 2006 to 2011. It would be interesting to study the dynamics of efficiency over long time horizon. If our main hypothesis is right, the proportion of clinical trials conducted by firms suppose to increase over time in the absence of external factors (ex: tax credit, or NIH funding of NPOs), or an isomorphism process between firms and NPOs might occur over time [Dimaggio and Powell (1983)]. Second, this study focus on duration of clinical trials, but other dimensions like cost and expected revenue could be integrated to our analysis if private data becomes publicly available. Third, this study didn't model or collect data about specific routines, practices, and procedures that confer competitive advantage to the firms in managing clinical trials. Neither, we show empirically the comparative advantage of NPOs in the early-stage research, what represents a big gap in the literature.

The policy implication of this study is why not publicly fund or provide tax incentives to firms to conduct clinical trials instead of NPOs? Our secondary results indicate that there is no difference in efficiency among firms in terms of scale and experience. Therefore, why doesn't federal agency fund small biotechs to run clinical trials instead of NPOs? Instead of funding clinical trials of NPOs, why not reallocate the funding to early-stage research such as basic science and research tools?

CHAPTER 3

THE IMPACT OF SCIENTIFIC AND TECHNOLOGICAL SPECIALIZATION ON THE TIMING OF TECHNOLOGICAL COLLABORATION IN THE BIOPHARMACEUTICAL INDUSTRY

3.1 Introduction

The diffusion of biopharmaceutical innovations is an important driver of increasing life span and social welfare seen since the Industrial Revolution [Reinganum (1989), Hoppe (2002), Lichtenberg (2005), Stoneman and Battisti (2010)]. QuintilesIMS (2016) predicts that the worldwide spending on drugs will increase from $1 trillion in 2014 to $1.5 trillion in 2021. Although the diffusion of innovation understood as technology embodied in the final product ready for commercialization is a classical topic studied in Economics [Griliches (1957), Berndt et al. (2003)], little is known about the diffusion of inventions or the timing of technological collaboration, a step that precedes the innovation. The exception are the studies of Katila and Mang (2003), Gans et al. (2008), Allain et al. (2009) that point out respectively how prior R&D collaboration experience, intellectual property rights protection, and competition in the final product speed up or delay the technological collaboration. Surprisingly, no empirical study investigated specifically how scientific and technological specialization would affect the timing of technological collaboration or research alliance.

An example of diffusion of invention is the messenger ribonucleic acid (mRNA) technology developed by Moderna Therapeutics. The mRNA technology consists in instructing the human cells to produce specific proteins against a disease. Moderna Therapeutics was founded in 2011. In 2012, this biotech firm already had important patents related to mRNA technology and scientific publica-

tions in journals like RNA and Nature Immunology. In 2013, AstraZeneca payed $240 million to Moderna Therapeutics for five years of collaboration in discovering and developing mRNA drugs against cancer and cardiometabolic diseases. From 2015 to 2016, Merck & Co. entered in multiple alliances with Moderna Therapeutics to acquire expertise in mRNA technology for cancer vaccine and infectious diseases.

From theoretical point of view, several models [Aghion and Tirole (1994), Aghion et al. (2008)] assume that technological specialization increases the innovation performance. However, there is no consensus on how scientific specialization affects the innovation performance. For some authors [Cohen and Levinthal (1989), Cockburn and Henderson (1998), Arora and Gambardella (1994)], scientific capabilities increase innovation performance via absorptive capacity, the ability to assimilate external knowledge and technology. For others [Nelson (1959), Arrow (1962), Partha and David (1994)], scientific knowledge is a public good and should be produced by non-profit organizations given the appropriation and underprovision problems. Aghion et al. (2008) defend that academic institutions have comparative advantage in the early stages of research (scientific domain); whereas firms have comparative advantage in the late stages of research (technological domain).

To test how scientific and technological specialization affect the timing of technological collaboration, we propose a new measure of specialization and timing of technological collaboration. As each alliance targets a specific technology (ex: monoclonal antibodies, recombinant DNA, stem cells, and etc) or disease (leukemia, diabetes, Parkinson's disease, and etc), we can count the scientific publications and patents that have these keywords in the title or abstract. Therefore, we measure how a firm is specialized in a specific technology or disease, and not in general categories such as anatomical therapeutic classes, scientific disciplines, or patent classes. Related to the timing of technological collaboration, we measure the timing between the licensor's first scientific publication (that cites a specific technology or diseases) and the starting date of alliance. We assume that the first scientific publication represents the "birth" of know-how or invention in a most elementary form. The drawback of this measure is that some firms engage in alliances, but they don't publish or might not have a publication that cites a specific technology or diseases. In this case, the timing of collaboration cannot be observed, but we correct for this selection problem, using the Heckman selection model.

After controlling for the factors described in the literature [Katila and Mang (2003), Gans et al.

(2008), Allain et al. (2009)] and other factors at alliance and firm level, we found that licensor specialization is the main factor to speed up the technological collaboration in the biopharmaceutical industry. We interpret this result as evidence that the licensor has a more critical role than the licensee in decreasing the uncertainty and asymmetric information problems around the market for technology. Only after the licensor signaling the existence of the technology and potential application without allowing others to invent around or reverse engineering, the licensee can decide to participate or not in the partnership.

Our result is important because the empirical literature of market for technology [Cassiman and Veugelers (2006), Higgins and Rodriguez (2006), Arora and Gambardella (2010), Ceccagnoli et al. (2014)] has focused on the determinants of demand side given the availability of data about the licensee (usually big firms), and we show evidence that supply side (usually small specialized firms) matters more. Based on our benchmark specification of the Heckman selection model, we estimated that an increase of 10% in the licensor scientific specialization speeds up the technological collaboration in about 4 months; whereas an increase of 10% in the licensor technological specialization speeds up the technological collaboration in about 5 months. Furthermore, our results suggest that licensee scientific and technological specialization in the same technology or disease of the licensor don't look to affect the timing of technological collaboration. A finding that is aligned with the theory of comparative advantage, as specialization must be in different tasks or stages, in order to explore the differences in opportunity cost and potential complementarities between activities.

3.2 Literature Review

Our research question is mainly related to two broad literature streams: 1) the diffusion of innovation; and 2) the innovation performance. In the first stream, we focus on the timing of technological collaboration, a topic not much explored. In the second stream, we cover specifically how scientific and technological specialization affects the innovation performance.

3.2.1 The Timing of Technological Collaboration

The landmark study concerned the diffusion of innovation is the paper of Griliches (1957), who analyzed how profitability affected the adoption of hybrid corn in United States from 1932 to 1956.

Since then, a vast theoretical and empirical literature emerged [Reinganum (1989), Hoppe (2002), Stoneman and Battisti (2010)], covering mainly how uncertainty, market structure, and strategic interaction affect the adoption of innovation. For example, Berndt et al. (2003) showed how the price, quality, entry timing, and consumption externality affect the diffusion rate of antiulcer drugs. Overall, this stream of literature focuses on innovation as a final product ready for commercialization, or an intermediate input or process ready to be employed. Little attention was given for the timing of technological collaboration, a step that precedes the innovation. The exception is Katila and Mang (2003), Gans et al. (2008), Allain et al. (2009) that analyzed how different factors might speed up or delay the technological collaboration.

Katila and Mang (2003) analyzed the history of 86 biopharmaceutical product-development projects from 1976 to1992 and found that R&D-intensive biotechnology firms that have prior R&D collaboration experience are faster in exploiting technological opportunities. They also found that the establishment of state biotechnology centers and an increasing of intellectual property protection speed up the technological collaboration. It is worth mentioning that Katila and Mang (2003) focus on exploitation and not on exploration of technological opportunities in the discovery phase. Neither they include different legal arrangements such as joint ventures, licensing, and co-licensing agreements in the sample.

Gans et al. (2008) investigated the impact of intellectual property system on the timing of licensing between start-up entrepreneurs and established firms across four industry sectors (biotechnology, electronics, software, and scientific instruments). In a sample of 198 technology licensing deals between 1990 and 1999, Gans et al. (2008) found that patent allowance (the administrative event when patent rights are clarified) is associated with a 70% increase in the hazard rate of achieving a licensing agreement. This finding suggests that uncertainty about intellectual property rights delays technological cooperation. However, other frictions in the market of technology due to asymmetric information, appropriability hazards, transaction costs might also play an important role on the timing of cooperation.

Allain et al. (2009) developed a model that predicts an inverted U-shaped relationship between competition and licensing delays in the case of asymmetric information about the value of technology. In a sample of 2212 licensing deals started since 1973 in the pharmaceutical industry, Allain et al. (2009) found empirical support that in concentrated markets, an increase in the number of

potential licensees delays licensing; whereas in competitive markets, an increase in the number of potential licensees speed up licensing. Although Allain et al. (2009) assume that the licensor is a perfect monopolist, there is also certain degree of competition in the supply side. In fact, the biopharmaceutical sector is characterized by far more small biotechnologies firms specialized in research than big consolidated pharmaceutical firms with comparative advantage in development and marketing of drugs.

3.2.2 Scientific and Technological Specialization

According with Arora and Gambardella (1994), scientific capability is the ability to evaluate information based on general and abstract principles; whereas technological capability is the ability to use or apply information. This distinction is relevant for the biopharmaceutical sector, as there is a labor division among academic institutions, small specialized firms, and big pharmaceutical firms. Stuart et al. (2007) argue that biotech firms evaluate and select scientific discoveries from universities, and pass it to big pharmaceutical firms.

Several studies [Henderson and Cockburn (1996), Cockburn and Henderson (1998), Nesta and Saviotti (2005), Garcia-Vega (2006), Miller (2006), Quintana-García and Benavides-Velasco (2008), Toh and Kim (2013)] investigate how technological specialization or diversification affects the innovation or firm performance. However, to the best of our knowledge no study investigates specifically how scientific specialization together with technological specialization affect the timing of technological collaboration.

From theoretical point of view, Aghion and Tirole (1994) point out that the innovation performance of an independent research unit (licensor) would be higher than a dependent research unit of the licensee. In the same line, based on the tradeoff between creative freedom vs pecuniary focus, Aghion et al. (2008) argue that the academic sector would have comparative advantage in the early stage of research (scientific domain); whereas the private sector would have comparative advantage in the late stages of research (technological domain).

Overall, the empirical literature found that technological diversification fosters the innovation and firm performance. The common theoretical justification is that economies of scope is more important than economies of scale at research level. The argument is that cross-fertilization among different technologies and the risk mitigation of pursuing parallel technological paths more than

compensate the cost of technological diversification. In the US. pharmaceutical sector from 1989 to 1997, Nesta and Saviotti (2005) found that a coherent technological diversification measured by patent classes enhances the firm's innovative performance measured by number of citation-weighted patents. In a set of European firms distributed in 15 different sectors from 1995 to 2000, Garcia-Vega (2006) also found that technological diversification enhances the firm's innovative performance measured by number of patents and R&D intensity. In a sample of U.S. biotechnology firms from 1976 to 2001, Quintana-García and Benavides-Velasco (2008) found that technological diversification has a stronger effect on radical innovation than incremental innovation. In a sample of U.S. firms from different industrial sectors from 1990, Miller (2006) found that technological diversification increases the market value of the firms.

However, contrary to the conventional wisdom that firms should hold real options across different technologies in face of uncertainty, Toh and Kim (2013) found that greater technological uncertainty increases the firm's technological specialization in the U.S. communications equipment industry. According to the Toh and Kim (2013), there is competition among technologies, and specialization fosters the probability of the firm's technology wins the competition.

3.3 Hypothesis

For the classical economists, the source of value creation is the division of labor. Under specific assumptions such as perfect information, zero transaction cost, and etc, if each part specializes in the task, in which he/she has comparative advantage, the total production of the goods will be maximized. Then, the parties can exchange the goods and realize the gains of specialization. The same logic applies to the strategic alliances. The idea behind of the existence of technological collaboration is that the parties can combine different capabilities and resources, in order to create value.

Besides value creation, licensor specialization in certain technology or disease also sends a stronger signal that can mitigate the market failure problems: I) information asymmetry [Akerlof (1970)], II) disclosure dilemma [Anton and Yao (1994, 2002)], and III) valuation difficulty [Contractor and Ra (2002)].

Usually the licensor has superior information about the technology. Specially when the tech-

nology has tacit elements embedded in the routines and employees of the firms [Nelson and Winter (1982)], the asymmetry tends to be high. Pisano (1997) argues that biotechs take advantage of asymmetric information and license-out the "lemons" compounds to pharmaceutical firms. However, Arora et al. (2009), Arora and Gambardella (2010) minimizes the importance of information asymmetric problems in the biopharmaceutical industry, alleging that: 1) both parties usually operate in the same industry or market; 2) the licensee is usually a sophisticated buyer with technical expertise similar to the licensor; and 3) regulatory authority imposes information disclosure in the clinical trials.

Related to the disclosure dilemma, the potential licensees want to learn more about how the technology can be integrated with their internal resources and capabilities, in order to create value. However, in the negotiation phase, the licensor might be reticent to disclosure information and reveal too much at the point that the potential licensees can proceed by their own, invent around, reverse-engineer, or simply misappropriate the technology. Arrow (1962) was the first to articulate the problem that the licensee has no incentive to pay the licensor after the information about the invention is disclosed.

Anton and Yao (1994, 2002) argue that this problem can be mitigated if there is competition among licensees or the invention can be partially disclosed. In the first case, the licensor can treat to destroy the rent of the first licensee, selling the invention to other potential licensees. In the second case, the licensor can partially signal the invention and receives the payment equivalent of the remaining undisclosed know-how. Overall, the market failure literature overlooks the learning aspect of strategic alliances and magnifies the problems of opportunistic behavior. The license agreement is designed to protect the proprietary know-how of both parties [Kale et al. (2000)]. For example, royalty clause, milestone payments, and equity purchase minimize the problem of misappropriation of intellectual property rights. Reputation and repetitive cooperation also inhibit opportunistic behavior.

Furthermore, the evaluation of the technology is not only difficulty because of the asymmetric information and disclosure dilemma. The technology per se might involve a high degree of complexity and uncertainty, making the monetary valuation a very complicate task. Even if the licensor wants to disclose all information available about the technology and tries to codify the tacit knowledge, the technology valuation is just a gross estimation with high variance given market and

commercial uncertainties. Therefore, the licensee willingness to pay is likely to be different than the reservation price of the licensor.

Considering that specialization creates value and sends a stronger signal about the scientific and technological capabilities of the licensor that mitigates some market failure problems, we hypothesize that:

Hypothesis 1a: Higher the scientific specialization of the licensor, faster is the technological collaboration.

Hypothesis 1b: Higher the technological specialization of the licensor, faster is the technological collaboration.

The logic of division of labor is that each part specializes in different tasks or stages given the differences in opportunity cost. If the licensor and licensee specialize in the same technology and disease, the probability of exploring complementarities between resources and capabilities decreases given the redundancy. Higher value surplus is created, when the licensee can access externally the technology that is unfeasible or expensive to develop internally. Therefore, if the licensee is an expert in the technology and disease domain of the licensor, the licensee has incentives to delay or even not starting the technology collaboration.

By the other hand, someone might argue that licensee specialization speeds up the technological collaboration based on the mitigation of the market failure problems described previously. Although the licensee specialization decreases the asymmetric information and uncertainty about the expected value of the technology, this impact is less critical if compared to the role of the licensor in revealing information and emitting the appropriate signal. Furthermore, to take advantage of external technology, the licensee might need absorptive capacity in different fields and not necessary be a specialist.

Therefore, as the licensee specialization is subject to two opposite effects described above, we hypothesize that:

Hypothesis 2a: The scientific specialization of the licensee in the same technology or disease of the licensor doesn't affect the timing of the technological collaboration.

Hypothesis 2b: The technological specialization of the licensee in the same technology or disease

of the licensor doesn't affect the timing of the technological collaboration.

3.4 Data

The main database is BiosciDB, a comprehensive peer-reviewed reference database on biopharma alliances updated by press releases, SEC filed contracts, and more than 5,000 unredacted contracts obtained through Freedom of Information Act (FOIA) disclosures. BiosciDB is a relatively new dataset and little explored if compared to other alliance datasets such as: Securities Data Company (SDC) and Recombinant Capital (RECAP) both from Thomson Reuters. SDC is the most famous and used dataset for empirical studies in alliances, but it doesn't have the same level of details of RECAP and BiosciDB related to biopharma alliances. The original creators of RECAP are the same creators of BiosciDB. Both databases follow the same structure with similar information. However, the creators of BiosciDB claimed to have substantially more unredacted contracts than RECAP. BiosciDB contains information about deal types, contract summary, parties (including tracking M&A), type of technology, target diseases, monetary compensation (size of the deal, royalties and equity), stage and date in which the contract was signed, and etc. One important functionality of BiosciDB website is that for almost all alliances, it is possible to download the original contract, a service not always offered by other alliance databases.

The patent dataset was extracted from Derwent Innovations Index in the Web of Science platform, a product of Thomson Reuters. Derwent Innovations Index covers over 14.3 million patents across 40 countries. Each patent record is corrected, analyzed, abstracted and manually indexed by subject experts from Thomson Reuters.

The scientific publication dataset was extracted from Scopus, a bibliographic database owned by Elsevier. Scopus covers over 21,500 academic journals and contains more than 60 million records. Each record contains detailed information about the scientific publication such as title, authors, address affiliation, keywords, abstract, citations, and etc.

The competition variable and ATC1 concordance (Anatomical Therapeutic Chemical Classification System) are from Pharmaprojects database, a product of Citeline. Pharmaprojects tracks drugs pipeline since 1980, covering all stages of drug development from discovery through clinical trials to drug approval and commercialization. Besides providing a detailed development history timeline

of over 60,000 drug profiles, Pharmaprojects includes licensing and sponsoring organization details together with a description of the disease, therapy, mechanism, delivery routes, chemical data and molecular structure.

Firm level accounting data (R&D expenditure, employment, assets, cash, etc.) was extracted from U.S. Compustat 1993–2013. About 17% of values were missing. We replaced these missing values using the Stata command "ipolate" (a linear interpolation method). As these variables change linearly over time and we observe the values for adjacent years by firm, we use years as predictor variable for the linear interpolation. All monetary variables were adjusted by inflation (base year 2013), using the Consumer Price Index from U.S. Bureau of Labor Statistics.

We started with 4384 alliances from BiosciDB. We removed alliances related to natural products or with missing values in the key variables such as technology and disease field. As the geographical focus of BiosciDB is United States, we also restricted the sample to publicly traded companies present in the U.S. Compustat, that is, nonprofit organizations as universities were removed from the sample. One limitation of dataset is that all information is limited to the alliance formation, that is, we don't observe the progression of alliance like success and failure rate.

Our final sample was composed by 845 alliances started from 1994 to 2014, and distributed among 357 licensors and 220 licensees.

3.4.1 Variable Definitions and Summary Statistics

Table 3.1 presents the summary statistics of the dependent, independent, and control variables described below. Table 3.2 presents the correlation matrix of these variables.

3.4.1.1 Dependent Variable

"Alliance Lag" is the main dependent variable; whereas "Pub.Likelihood" is the dependent variable of the selection model.

We defined the dependent variable "Alliance Lag" measured in years to be equal to the "Alliance Date" minus the "First Publication Date". The "Alliance Date" is the starting date of the alliance a. The "First Publication Date" is the date of the oldest publication of the licensor s that cites a specific technology t or a targeted diseases d. We restricted our search for scientific publications to 25 years

Table 3.1: Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
Alliance Lag	679	9.57	7.94	0	25
Pub.Likelihood	845	0.80	0.40	0	1
Licensor Sci. Spec.	845	0.25	0.29	0	1
Licensor Tech. Spec.	845	0.30	0.29	0	1
Licensee Sci. Spec.	845	0.15	0.21	0	1
Licensee Tech. Spec.	845	0.17	0.21	0	1
Licensor Patent Classes	845	9.20	10.4	0	98
Licensee Patent Classes	845	13.9	18.7	0	249
Licensor Experience	845	4.35	5.21	0	29
Licensee Experience	845	17.1	21.3	0	101
Prior Collaboration	845	0.18	0.51	0	5
Technology Complexity	845	1.25	0.68	0	2
Competition	845	0.77	0.67	0	1.62
Equity Alliance	845	0.31	0.46	0	1
Licensor Cash ($Billion)	845	0.83	2.29	-1.26	15.6
Licensee Cash ($Billion)	845	2.25	3.43	-1.38	21.1
Licensor R&D Intensity	845	0.59	1.20	0	5.84
Licensee R&D Intensity	845	0.27	0.57	0	3.44
Licensor Size	845	11.5	27.5	0.0010	138
Licensee Size	845	34.5	40.4	0.0010	333

before the "Alliance Date". The results of Table 3.3 and 3.4 remain the same if we restrict the search for publications to 20 or 30 years before the "Alliance Date". The "First Publication Date" is a proxy for the "birth" of know-how or technology in a most elementary form. We don't use the "First Patent Date", because patent represents a consolidated technology, and several alliances precede the patent application or grant [Gans et al. (2008)]. It takes on average 10 years from the "birth" of know-how or technology until the technological cooperation.

The other dependent variable "Pub.Likelihood" is coded as 1 if the licensor s published any academic article 25 years before the "Alliance Date" that cites a specific technology t or a targeted diseases d, and 0 otherwise. For 80% of alliances, the licensor published at least one academic paper related to technology t or a targeted diseases d. By definition "Alliance Lag" only can be observed if "Pub. Likelihood" is equal to 1. Among 845 alliances, we observe 679 alliance lags. In order to deal with this 166 missing values for "Alliance Lag", we use the variable "Pub.Likelihood" as dependent variable in the selection model described in the next section. As the variable "Pub.Likelihood" doesn't use all the information available about scientific publications, we

Table 3.2: Correlation Matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Alliance Lag	(1)	1.00													
Pub.Likelihood	(2)	0.00	1.00												
Licensor Sci. Spec.	(3)	-0.29**	0.44**	1.00											
Licensor Tech. Spec.	(4)	-0.19**	0.10**	0.44**	1.00										
Licensee Sci. Spec.	(5)	0.06	0.08*	0.10**	0.09**	1.00									
Licensee Tech. Spec.	(6)	0.10**	0.12**	0.11**	0.12**	0.54**	1.00								
Licensor Experience	(7)	0.44**	0.21**	-0.08*	-0.07*	0.02	0.11**	1.00							
Licensee Experience	(8)	-0.03	0.10**	0.19**	0.23**	-0.04	0.03	0.02	1.00						
Prior Collaboration	(9)	0.06	0.01	0.00	0.07*	-0.02	0.01	0.25**	0.15**	1.00					
Technology Complexity	(10)	0.01	0.05	0.17**	0.15**	0.03	0.07*	0.04	0.11**	0.08*	1.00				
Competition	(11)	-0.04	0.07	0.11**	0.10**	0.10**	0.10**	0.05	-0.01	0.03	0.12**	1.00			
Equity Alliance	(12)	-0.07	-0.00	-0.02	-0.03	-0.06	-0.00	-0.04	-0.08*	-0.04	0.07*	0.02	1.00		
Licensor Patent Classes	(13)	0.50**	0.28**	0.10**	0.14**	0.17**	0.18**	0.35**	0.07	0.12**	0.00	-0.00	-0.05	1.00	
Licensee Patent Classes	(14)	-0.07	0.06	0.14**	0.22**	0.19**	0.27**	-0.02	0.29**	0.07*	0.05	0.07*	-0.04	0.04	1.00

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

created the variable "# Pub.Citation" used as dependent variable of the negative binomial model of Table 3.3. "# Pub.Citation" is the number of scientific publications of the licensor s that cites a specific technology t or a targeted diseases d. Each publication was weighted by the number of citations received until August 2016.

3.4.1.2 Independent Variables

"Licensor Sci. Spec.". Licensor scientific specialization is the proportion of the scientific publication of the licensor s that cites a specific technology t or a targeted diseases d during 25 years before the "Alliance Date". Each publication was weighed by citations received until August 2016, in order to control for the impact and quality of the signal emitted by the licensor s. We don't use the raw number of publications that cites a specific technology t or a targeted diseases d, because this measure would capture size effect or overall scientific capability instead of specialization. For example, normalized measure of innovation like scientific publications and patents divided by R&D expenditure indicate that biotech firms are more innovative than big pharmaceutical firms [Arora et al. (2014)].

"Licensor Tech. Spec.". Licensor technological specialization is the proportion of patents of the licensor s that cites a specific technology t or a targeted diseases d during 25 years before the "Alliance Date". Our measure of technological specialization is more specific than employed in the literature. For example, as a proxy for technological specialization, Toh and Kim (2013) use the Herfindahl index (a measure of concentration) calculated based on patent classes.

Unfortunately, it is not possible to import from Derwent Innovations Index the number of citations that a patent received. Thomson Reuters might add this functionality in the future. However, we weighted each patent in the following way: a triadic patent filled at the European Patent Office (EPO), the United States Patent and Trademark Office (USPTO) and the Japan Patent Office (JPO) receives value 3; whereas a non-triadic patent receives value 1.

"Licensee Sci. Spec.". Licensee scientific specialization is the proportion of the scientific publication of the licensee b that cites a specific technology t or a targeted diseases d during 25 years before the "Alliance Date". Each publication was weighed by citations received until August 2016, in order to control for the strength of scientific capability of the licensee b.

"Licensee Tech. Spec.". Licensee technological specialization is the proportion of patents of

the licensee b that cites a specific technology t or a targeted diseases d during 25 years before the "Alliance Date". We also weighted each patent using the triadic patent method.

It is worth mentioning that the weight methods used for the construction of the independent variables above are common employed in the literature, in order to deal with noise and uncertainty about the value of each publication and patent. However, in this paper the use or not use of these weight methods doesn't affect the results of Table 3.3 and 3.4, what makes sense considering that we already restricted the publications and patents to specific technology t or a targeted diseases d.

"Licensor Patent Classes". Given the patents of licensor s that cite a specific technology t or a targeted diseases d during the 25 years before the "Alliance Date", we sum up the total number of 3-digit patent classes. Derwent Innovations Index divides the patents in three groups: chemical, engineering, and electrical. One division of chemical is "B - Pharmaceuticals", in which is composed by 3-digit classes, such as: B01 - Steroids; B02 - Fused ring heterocyclics; and etc. The total numbers of 3-digit patent classes is 291 and the classes are not mutually exclusive.

"Licensee Patent Classes". Given the patents of licensee b that cite a specific technology t or a targeted diseases d during the 25 years before the "Alliance Date", we sum up the total number of 3-digit patent classes.

The "Licensor Patent Classes" measures the technological diversification given a specific technology t or a targeted diseases d; whereas the variable "Licensor Tech. Spec." measures the licensor technological specialization in relation to his whole portfolio of patents. Furthermore, except the "Licensor Sci. Spec." and "Licensor Sci. Spec." that is moderate correlated (0.44), all independent variables are weakly pairwise correlated according with Table 3.2. The results remain the same in Tables 3.3 and 3.4, if we drop the variables "Licensor Patent Classes" and "Licensee Patent Classes".

3.4.1.3 Control Variables

"Licensor Experience". It is the cumulative number of alliances engaged by the licensor s until the starting date of the alliance a in the BioSciDB since 1980.

"Licensee Experience". It is the cumulative number of alliances engaged by the licensee b until the starting date of the alliance a in the BioSciDB since 1980. The average number of previous alliance is about 4 for the licensor and 17 for the licensee.

"Prior Collaboration". It is the cumulative number of alliances in the BioSciDB since 1980, in which the licensor s and the licensee b cooperate before the starting date of the alliance a. The mean for "Prior Collaboration" is 0.18. This lower value indicates that few firms engage in alliances with the same partner. One explanation is that many licensors are acquired by the licensee after the technological collaboration [Higgins and Rodriguez (2006)]. Katila and Mang (2003) found that previous R&D collaboration with the same partner accelerates the actual technological collaboration.

"Technology Complexity". We classify each technology t of alliance a, in three groups. Less complex technologies related to drug delivery and diagnostics were assigned value 0. Moderate complex technologies (ex: screening and chemical synthesis) related to chemical compounds were assigned value 1. More sophisticated technologies related to biologic compounds (ex: monoclonal antibodies, recombinant DNA, and stem cells) were assigned value 2. Each category represents respectively 12.2%, 49.2%, and 38.6% of the sample.

"Competition". It is the cumulative number of drug approval for a targeted diseases d by regulatory authorities around the world since 1977 until the starting date of the alliance a. We divided this variable by 100. The mean of this variable is 0.77. This number is relatively big because several alliances target many diseases or a set of connected diseases. In the chemical industry, Fosfuri (2006) found that the number of technology licensing follows an inverted U-shaped relationship with the number of potential licensors.

"Equity Alliance". This variable was coded as 1, if the licensor s cedes any degree of control or ownership to the licensee b, and 0 otherwise. The literature of governance structure [Pisano (1989), Oxley (1997), Das and Teng (2000), Santoro and McGill (2005)] points out that equity alliances are more likely to involve a "learning process" than mere transfer or access of technology.

"Licensor Cash". It is the amount of any immediately negotiable medium of exchange of the licensor s in the year before the starting date of the alliance a. This variable was measured in billions of US dollars. The idea is to control for firm liquidity as in Katila and Mang (2003), Toh and Kim (2013). Under financial constraint the licensor might try to speed up any technological collaboration.

"Licensee Cash". It is the amount of any immediately negotiable medium of exchange of the licensee b in the year before the starting date of the alliance a. This variable was measured in billions

of US dollars. If the licensee has abundance of cash given higher drug sales or business divestiture, the licensee might be interested to speed up the technological collaboration. On average, the licensee (U$ 2.25 billion) has about 3 times more cash than the licensor (U$ 830 million).

"Licensor R&D Intensity". It is ratio between the R&D expenditure and total assets of the licensor s in the year before the starting date of the alliance a. The mean of licensor R&D intensity (0.59) is about twice higher than the mean of licensee R&D intensity (0.27). As the licensor is the "seller" of R&D and the licensee is the "buyer" of R&D, this R&D expenditure difference might represent the optimum level for each side. Concerned the timing of technological collaboration, Katila and Mang (2003) found that the R&D intensity of the licensor speeds up the alliance formation.

"Licensee R&D Intensity". It is ratio between the R&D expenditure and total assets of the licensee b in the year before the starting date of the alliance a. Usually R&D intensity is used as a proxy for absorptive capacity of the licensee [Cohen and Levinthal (1989), Arora and Gambardella (1990)]. However, higher R&D intensity is not only related to higher capacity to identify and use external technology, but also with higher internal capabilities to develop the technology in-house as well. There is no consensus in the literature if internal R&D and external R&D (alliances) are substitutes or complements [Colombo and Garrone (1996), Cassiman and Veugelers (2006), Ceccagnoli et al. (2014)]. Less is known how the licensee R&D intensity affects the timing of technological collaboration.

"Licensor Size". Number of employees of licensor s in the year before the starting date of the alliance a. This variable was divided by 1000.

"Licensee Size". Number of employees of licensee b in the year before the starting date of the alliance a. This variable was divided by 1000. On average the licensee is three times bigger than the licensor. Quite often the licensee, the buyer of technology, is a consolidated big pharmaceutical firm; whereas the licensor, the seller of technology, is a small biotech.

"ATC1". They are dummies for 12 ATC level 1 not mutually exclusive classes: ATC_A (Alimentary tract and metabolism), ATC_B (Blood and blood forming organs), ATC_C (Cardiovascular system), ATC_D (Dermatologicals), ATC_G (Genito-urinary system and sex hormones), ATC_H (Systemic hormonal preparations), ATC_J (Antiinfectives for systemic use), ATC_L (Antineoplastic and immunomodulating agents), ATC_M (Musculo-skeletal system), ATC_N (Nervous system), ATC_R (Respiratory system), and ATC_S (Sensory organs). Based on diseases d and Pharmapro-

jects concordance, each alliance a was classified in one or more ATC1, and there is no alliance covering the ATC_P (Antiparasitic products) in the final sample.

"Year". Dummies for alliance starting year from 1994 to 2014.

3.5 Empirical Strategy

Our goal is to estimate how scientific and technological specialization impact the timing of the technological collaboration. For that purpose, we compiled a unique dataset at alliance level. As we observe for each alliance the type of technology employed (ex: monoclonal antibodies, recombinant DNA, stem cells, and etc) and the targeted disease (leukemia, diabetes, Parkinson's disease, and etc), we could search the specific patents and publications that cover both technology and disease. This approach represents a methodological advance, as the literature has been using the absolute number of patents and publication at firm level, and we were able to delimit the patents and publication at alliance level. For example, in 2016 the accumulated patent portfolio of Merck was above 10,900 patents, but only 23 patents were related to monoclonal antibodies technology.

The dependent variable "Alliance Lag" that represents the timing of technological collaboration can be observed only if the licensor s decides to publish. Firms have different propensity to publish based on strategic reasons. For example, in the case of recombinant DNA technology, Figure 3.1 and 3.2 show the top 10 firms based respectively on publications and patents. Big pharmaceutical firms like Pfizer and Merck are top 10 in terms of scientific publications but not in terms of patents in recombinant DNA. Two Japanese firms (Kyowa Kirin and Takeda) are in the second and third position in terms of patents in recombinant DNA, but they don't have much scientific publications compared to American and European firms. For some firms, publishing and patenting are complements; whereas for others substitutes. Probably, firms that don't publish are more likely to rely on patents for signaling purpose. If we run only one separated regression to estimate the "Alliance Lag", the coefficients have to be interpreted as conditioned to firms that decided to publish. In order to generalize the results to the larger sample, sample selection models [Heckman (1976), Puhani (2000), Cameron and Trivedi (2005)] can be used such as the limited information maximum likelihood (LIML) method (known as Heckman selection model) or the full information maximum likelihood (FIML) estimator.

Figure 3.1: Top 10 Firms based on Scientific Publication in Recombinant DNA

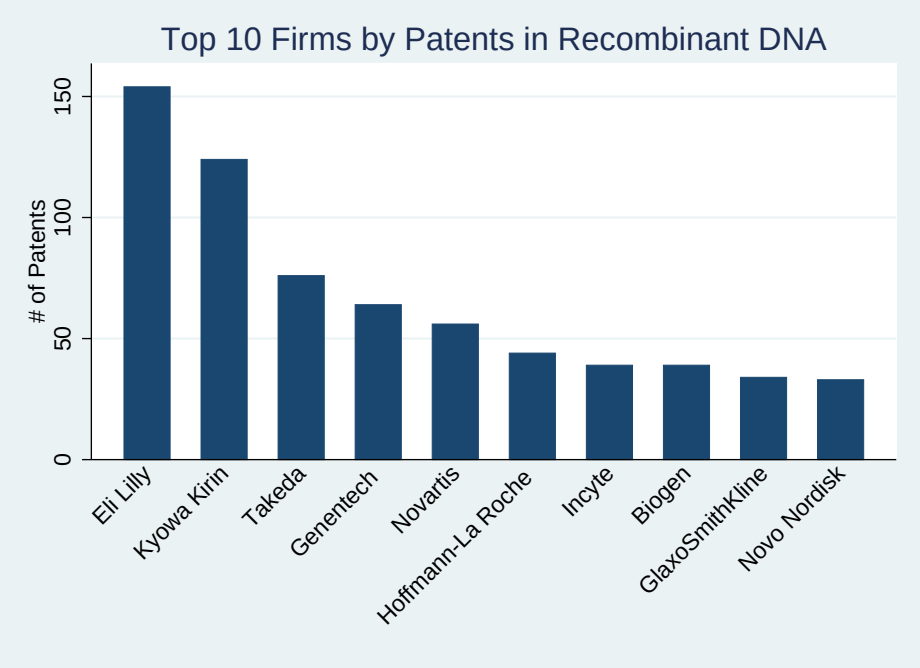


Figure 3.2: Top 10 Firms based on Patents in Recombinant DNA

The selection models consist of selection equation and outcome equation:

$$Pr[Pub.Likelihood]_1 = x_1^{'}\beta_1 + u_1 \tag{3.1}$$

$$Alliance\_Lag_2 = x_2^{'}\beta_2 + v_2 \tag{3.2}$$

where $Pr[Pub.Likelihood]_1$ is a dummy variable that captures the probability of the licensor s to publish, $Alliance\_Lag_2$ is the years between the "birth of the technology" and date of alliance, $x_1^{'}$ and $x_2^{'}$ are the vectors of explanatory variables used respectively in the selection and outcome equations, β_1 and β_2 are the vectors of coefficients to be estimated, and u_1 and v_2 are the usual error terms for the correspondent selection and outcome equations.

The Heckman selection method consists first in estimating the selection equation 3.1 using the Probit regression. Then, the predicted probabilities generated in the first step is used to construct the sample correction variable λ (the inverse of the Mills' ratio):

$$\lambda = \phi(x_1^{'}\hat{\beta}_1)/\Phi(x_1^{'}\hat{\beta}_1) \tag{3.3}$$

where $\phi(\cdot)$ is the normal probability density function, $\Phi(\cdot)$ is the normal cumulative distribution function, and $\hat{\beta}_1$ is the vector of coefficients obtained from the first step Probit regression. In the second step, the outcome equation 3.2 is estimated by OLS, using the λ as a control variable:

$$Alliance\_Lag_2 = \alpha_2 Licensor\_Sci.Spec. + \delta_2 Licensor\_Tech.Spec.$$

$$+\eta_2 Licensee\_Sci.Spec. + \theta_2 Licensee\_Tech.Spec. + x_2^{'}\beta_2 + \sigma_{12}\lambda_1 + v_2 \tag{3.4}$$

where $Licensor\_Sci.\ Spec.$, $Licensor\_Tech.Spec.$, $Licensee\_Sci.Spec.$, and $Licensee\_Tech$ $.Spec.$ are the variables of interest, α_2, δ_2, η_2, and θ_2 are the respective coefficients of the variables of interest, σ_{12} is the error covariance between the first step equation and second step equation. The Wald test can be used to test if the errors are uncorrelated ($\sigma_{12} = 0$). If the errors are correlated ($\sigma_{12} \neq 0$), the sample selection correction is necessary. Without the correction, the OLS estimation

would suffer from the omitted variable bias, leading to inconsistent estimation of α_2, δ_2, η_2, and θ_2. The standard errors are adjusted according with Heckman (1979), considering the first step and the estimation error of λ.

Instead of using the two steps of Heckman selection method, it is possible to estimate the selection equation 3.1 and outcome equation 3.2 by using the full information maximum likelihood estimator with robust standard errors clustered by licensee level. Although Monte Carlo studies [Puhani (2000)] show that FIML estimator are in general more efficient than the Heckman selection estimator, the FIML estimator demands the stronger assumption that the correlated errors (u_1 and v_2) are joint normally distributed and homoskedastic [Cameron and Trivedi (2005)].

Besides adjusting for selection, we added a set of control variables at alliance and firm level described in the previous section to equation 3.2. The idea is to control for any potential factor described in the literature that might affect the timing of collaboration, and is correlated with our independent variable "specialization", avoiding any omitted variable problem. We also added 12 ATC1 dummies variables to control for opportunity technology or specificities in different anatomical therapeutic classes. Finally, we added dummy variables for years to control for potential shocks or trends over the time. Given the relatively small sample size, we didn't add dummies for the licensors and licensees. Otherwise, we would have more variables than observations. However, we tested different specifications with dummies for 10, 25, and 50 big licensors and licensees based on numbers of alliances, and the results of Tables 3.3 and 3.4 remained the same. We also tested interaction terms among the dependent variables, but they were not statistically significant.

3.6 Results

Table 3.3 presents the estimation of selection equation. All regressions control for ATC1 (12 anatomical therapeutic classes) and starting year of the alliance. Robust standard errors are clustered at licensor level in the regression 1; whereas for regressions 2 and 3, standard errors were calculated and adjusted jointly with Alliance Lag equations of Table 3.4 . In the regression 1, a Negative Binomial estimation of the number of publications weighed by citations is presented. The logic of regression 1 is to use the full information about scientific publication to justify a set of control variables to the selection regressions 2 and 3. The over-dispersion parameter α is equal 3.7.

Table 3.3: Estimation of Selection Equation

	Neg. Binomial	Heckman	FIML
	(1)	(2)	(3)
	# Pub.Citation	Pub.Likelihood	Pub.Likelihood
Licensor Tech. Spec.	0.58	0.09	0.04
	(0.35)	(0.22)	(0.21)
Licensor Patent Classes	0.08***	0.09***	0.09***
	(0.01)	(0.01)	(0.02)
Licensor Experience	0.12***	0.02	0.02
	(0.02)	(0.02)	(0.02)
Technology Complexity	0.76***	0.09	0.08
	(0.14)	(0.09)	(0.07)
Competition	0.00	0.05	0.02
	(0.10)	(0.09)	(0.09)
Licensor Cash	0.13***	0.12	0.21
	(0.04)	(0.12)	(0.18)
Licensor R&D Intensity	-0.04	0.00	0.01
	(0.07)	(0.04)	(0.04)
Licensor Size	0.03***	0.01*	0.02*
	(0.00)	(0.01)	(0.01)
ATC1 (12 Dummies)	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
α	3.70		
Log Likelihood	-6563.34	-321.10	-2431.39
Observations	845	845	845

Note: For model 1, robust standard erros are in parentheses (clustered by licensor). For models 2 and 3, Stata calculated and adjusted the standard erros jointly with Alliance Lag equations of Table 4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Based on likelihood-ratio test, we reject the null hypothesis that α is equal 0 at 1% level of significance, indicating that Negative Binomial regression, in which the conditional variance exceeds the conditional mean, is more appropriate than the Poisson regression. Regression 1 suggests that big licensors with more experience and cash publish more papers with higher impact that cites a specific technology t or a targeted diseases d. Technological diversification measured by patent classes, and technology complexity are also positive associated with higher number of publications.

Regressions 2 and 3 of Table 3.3 are respectively the Heckman and FIML selection equations for the outcome equations 4 and 5 of Table 3.4. Both regressions, a Probit estimation of "Publication Likelihood" of the licensor, present similar results. The selection equations have less regressors than the outcome equations, because we assume that variables related to the licensee don't affect the publishing probability of the licensor. Although the variable "Licensor Tech. Spec." is not statistically

significant, the significance of the variable "Licensor Patent Classes" indicates that patenting in different fields increases the probability of publication. Other variable associated with the probability of publication is the "Licensor Size". It might be the case that small firms have lower probability to publish, because they need to send a stronger signal via patent instead of publication. Overall, the selection regressions 2 and 3 indicate that selection problem is that relatively small firms with lower scope (experience in different technical fields) is less likely to be in the sample of Table 3.4.

Table 3.4 presents the estimation of Alliance Lag equation. Regressions 1, 2, 3, estimate the "Alliance Lag" by OLS independent of the selection equation; whereas regressions 4 and 5 estimate "Alliance Lag" with the sample selection correction λ respectively by Heckman and FIML estimators. All regressions control for ATC1 (12 anatomical therapeutic classes) and starting year of the alliance. For models 1, 2, and 3, robust standard errors are clustered by licensee; whereas for models 4, and 5, the standard errors were calculated, considering the estimation error of λ (the sample selection correction variable). The results provide empirical support for H1 and H2. Related to H1 that the licensor scientific and technological specialization speeds up the technological collaboration. For all regressions, the coefficients of "Licensor Sci. Spec." and "Licensor Tech. Spec." are negative and statistically significant at 1%. Overall, each time that a set of control variables is added from regression 1 to 3, the magnitude in absolute terms of the coefficients of "Licensor Sci. Spec." and "Licensor Tech. Spec." decreases as part of variation is removed. Concerned H2 that the licensee specialization doesn't affect the timing of technological collaboration, the coefficients of "Licensee Sci. Spec." and "Licensor Tech. Spec." are not statistically significant at even 10%. Other evidences that supports H1 and H2 are respectively the positive and statistically significant coefficient of "Licensor Patent Classes" and the non-statistically significant coefficient of "Licensee Patent Classes". The variable "Licensor Patent Classes" indicates that an increase of technological diversification measured by patent classes delays the technological collaboration; whereas the variable "Licensee Patent Classes" has no effect on the collaboration timing.

In regressions 4 and 5 of Table 3.4, the sample correction λ is statistically significant different from zero at any p-value higher than 1%. Therefore, we cannot reject that the sample selection correction is necessary. A pattern that can be noticed is that the magnitude in absolute terms of "Licensor Sci. Spec." from regressions 1 to 3 slightly decreases when compared to regressions 4 and 5. However, in the case of the variable "Licensor Tech. Spec." the opposite occurs, that is, the

Table 3.4: Estimation of Alliance Lag Equation

	(1) OLS	(2) OLS	(3) OLS	(4) Heckman	(5) FIML
Licensor Sci. Spec.	-7.30***	-5.43***	-4.23***	-3.84***	-3.94***
	(1.13)	(0.98)	(0.93)	(1.38)	(0.81)
Licensor Tech. Spec.	-3.26***	-3.61***	-2.57***	-3.97**	-3.24***
	(1.12)	(0.92)	(0.82)	(1.67)	(0.85)
Licensee Sci. Spec.	1.51	-0.17	-0.39	0.01	0.22
	(1.86)	(1.50)	(1.45)	(2.08)	(1.25)
Licensee Tech. Spec.	1.21	0.36	-0.75	-1.21	-1.07
	(1.83)	(1.52)	(1.58)	(2.11)	(1.31)
Licensor Patent Classes		0.29***	0.24***	0.16***	0.21***
		(0.03)	(0.03)	(0.05)	(0.03)
Licensee Patent Classes		-0.01	0.02	0.01	0.01
		(0.01)	(0.01)	(0.02)	(0.01)
Licensor Experience		0.32***	0.24***	0.16*	0.21***
		(0.05)	(0.06)	(0.09)	(0.05)
Licensee Experience		-0.01	0.01	0.01	0.01
		(0.01)	(0.01)	(0.02)	(0.01)
Prior Collaboration		-0.31	-0.17	-0.19	-0.15
		(0.50)	(0.51)	(0.74)	(0.50)
Technology Complexity		0.79*	0.97**	0.87	0.92**
		(0.41)	(0.40)	(0.61)	(0.39)
Competition		-0.36	-0.16	-0.23	-0.23
		(0.35)	(0.34)	(0.61)	(0.35)
Equity Alliance		-0.61	-0.52	-0.40	-0.44
		(0.53)	(0.51)	(0.75)	(0.47)
Licensor Cash			0.38**	0.37	0.36**
			(0.16)	(0.24)	(0.15)
Licensee Cash			0.05	0.06	0.03
			(0.08)	(0.14)	(0.07)
Licensor R&D Intensity			-0.32*	-0.17	-0.26
			(0.18)	(0.35)	(0.19)
Licensee R&D Intensity			1.41**	1.22*	1.42***
			(0.57)	(0.73)	(0.52)
Licensor Size			0.02	-0.00	0.01
			(0.02)	(0.02)	(0.02)
Licensee Size			-0.02**	-0.01	-0.01*
			(0.01)	(0.01)	(0.01)
λ				-10.19***	
				(2.65)	
ATC1 (12 Dummies)	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
R^2	0.23	0.46	0.50		
Log Likelihood	-2280.76	-2159.55	-2135.51		-2431.39
Observations	679	679	679	845	845

Note: The variable λ (the inverse of the Mills' ratio) is an adjustment for sample selection.
For models 1, 2, and 3, robust standard errors are in parentheses (clustered by licensee).
For models 4 and 5, the standard errors in parentheses were calculated, considering the estimation error of λ.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

magnitude in absolute terms of "Licensor Patent" from regressions 4 to 5 increases to some extent if compared to regressions 1 to 3. Our interpretation is that without sample correction, regression 1 to 3 overestimate the impact of "Licensor Sci. Spec." and underestimate the impact of "Licensor Tech. Spec.", because firms that don't publish are more likely to rely on patenting for signaling purpose. Our preferred specification the Heckman Selection model, in which assumes less stronger assumption about the error term, indicates that the impact of "Licensor Sci. Spec." and "Licensor Tech. Spec." are equally relevant. Based on regression 4, we estimate that an increase of 10% in the proportion of publication that cites a specific technology t or a targeted diseases d speeds up the technological collaboration in about 4 months; whereas an increase of 10% in the proportion of patent that cites a specific technology t or a targeted diseases d speeds up the technological collaboration in about 5 months.

We are cautions to interpret the coefficients of control variables, because the most part are not statistically significant and stable across different specifications. Therefore, we are going to focus on the two variables that are statistically significant in the regression 4 and 5 of Table 3.4: "Licensor Experience" and "Licensee R&D Intensity". An increase of "Licensor Experience" is associated with a delay on technological collaboration. One explanation is that relatively older and bigger licensors with higher experience in previous alliances prefer to cooperate late, in order to decrease the uncertainty about the value of technology and to pressure for more beneficial division of the surplus. Based on data from BiosciDB, we noticed that the licensor receives less royalties when the alliance is signed in the early stage of drug development than late stages. Concerned to the "Licensee R&D Intensity", this variable is also associated with an increase of the collaboration timing. Some authors [Katila and Mang (2003)] use R&D intensity as a proxy for absorptive capacity. Although more R&D expenditure increases the capacity of the firm to identify, select, and use external technology, it is also increases his own in-house capabilities to develop the technology internally. The positive sign of the coefficient "Licensee R&D Intensity" suggests that the second effect dominates the first one.

3.7 Conclusion

After controlling for several factors at alliance and firm level, and selection issue, we showed that licensor specialization speeds up the timing of technological collaboration. Based on our preferred specification, we estimated that an increase of 10% in the scientific specialization speeds up the technological collaboration in about 4 months; whereas an increase of 10% in the technological specialization speeds up the technological collaboration in about 5 months. Furthermore, we found that both licensee scientific specialization and licensee technological specialization in the same technology or disease of the licensor don't affect the timing of the technological collaboration. We interpret these results as evidence that licensor has an important role in revealing information and emitting the appropriate signal to decrease the problems of asymmetric information and uncertainty around the market for technology.

Therefore, if policy makers want to accelerate the technological collaboration, licensor specialization should be incentivized instead of licensee specialization, avoiding the duplicity of resources and exploring the comparative advantage of each party. For example, a R&D tax credit that targets and concedes benefits to relatively small-medium specialist firms would have higher impact in the total social welfare than an equalitarian policy that benefits large pharmaceutical firms that usually are the buyer of the technology.

Other takeaway of this study is that licensor scientific specialization matters in similar way to the licensor technological specialization. As biopharmaceutical sector is a very science-based sector, the labor division in scientific domain vs technological domain doesn't match well with the typology: nonprofit organizations vs firms. Therefore, a policy that assumes that scientific knowledge is a public good and should be exclusively produced or allocated to nonprofit organizations would probably delay the technological collaboration. Bibliometric studies [Koenig (1983), Rafols et al. (2014)] indicate that scientific papers of pharmaceutical firms in basic biomedical research (ex: pharmacology, biochemistry, molecular biology, and etc) are highly cited as research supported by NIH (U.S. National Institutes of Health).

To better inform R&D policies, it would be important to investigate and to model the determinants of scientific and technological specialization for example: environmental conditions at organization level or institutional factors at country level. Given the data restriction, this study focused on

alliance formation in United States. It would be interesting to follow up and check the success rate of alliances in other countries. Other interesting future research would be to test the same methodology of restricting the scientific publications and patents that cite a specific technology to other high tech sectors as software and semiconductor.

CHAPTER 4

INNOVATION PRODUCTIVITY OF MAKING, ALLYING, AND BUYING IN THE PHARMACEUTICAL INDUSTRY

4.1 Introduction

R&D portfolio of a pharmaceutical firm can be divided in three categories: 1) internal projects; 2) projects developed in alliance with other firms; and 3) projects originated from M&A. The literature [Pisano (1997), Danzon et al. (2005), Guedj (2005), Arora et al. (2009)] compared the innovative performance of categories 1 and 2 under the dichotomy: make vs buy, or internal vs external R&D projects. Somehow, the previous literature diluted category 3 in categories 1 and 2. However, a snapshot of R&D portfolio of medium and big firms reveals that substantial part of R&D projects comes from M&A. In fact, to bring new drugs to the market, managers have to consider the trade-off among the three strategies: 1) to develop internally the R&D projects; 2) to develop the R&D projects in cooperation with other firms, for example in-license the compound; and 3) to acquire a bundle of R&D projects via M&A. For example, Novartis and GlaxoSmithKline exchange assets in 2014 to strengthen their area of expertise. Novartis acquired the oncology division of GlaxoSmithKline for $16 billion. In exchange, GlaxoSmithKline acquired the vaccines division from Novartis for $7.1 billion.

Little attention was given to category 3, the exception was the finding of Grabowski and Kyle (2008) that indicates that post-merger projects have higher probability of success than pre-merger projects. Differently from Grabowski and Kyle (2008) that compared the R&D projects before and after the M&A, the goal of this study is to compare the innovation performance at intra-project level based on three different sources: internal, alliance, and M&A.

Furthermore, the literature [Pisano (1997), Danzon et al. (2005), Guedj (2005), Arora et al. (2009)] provides a mixed evidence about the impact of alliances on innovation productivity. Pisano (1997), Guedj (2005) found that collaborative projects have higher failure rate than vertically integrated projects in the pharmaceutical industry. Specifically, Guedj (2005) found that collaborative projects have higher success rate in moving from Phase I to Phase II, but lower success rate in moving from Phase II to Phase III and in receiving FDA approval. The explanation of Pisano (1997) is that Biotechs develop internally the best compounds and license-out the "lemons" compounds. In the same line of reasoning, Guedj (2005) complements that Biotechs advances low quality projects in the early clinical trials, in order to avoid a negative signal that would jeopardize the access of funding. However, Danzon et al. (2005) found that in-licensed compounds have higher success rate than in-house compounds. According to the Danzon et al. (2005), licensee's experience in drug development increases the success rate, specially the phase III, which is more complex and expensive. Arora et al. (2009) found that in-licensed compounds have the same probability of success as in-house compounds. According to the Arora et al. (2009), comparing to big Pharmas, Biotech has lower probability to receive FDA approval, not because of "lemons" problem, but because Biotechs have on average lower quality compounds.

One main contribution of this study is to rationalize the conflicting findings in the previous literature about innovation productivity of internal R&D projects vs external R&D projects, controlling explicitly for M&A. This study found that if external R&D projects are defined as projects developed in research alliance and projects originated from M&A, there is no difference between the probability of success of internal R&D projects and external R&D projects. However, when the external R&D projects is split in a relevant category research alliance and M&A, R&D projects developed by research alliance have higher probability of success than internal R&D projects, but R&D projects originated from M&A have lower probability of success than internal R&D projects. This study also found that the probability of success of research alliance and M&A depends positively on who is the partner or the targeted firm. Research alliances between big pharmaceutical firms have higher probability of success than research alliances, in which one part is a small or medium firm. In the same line, R&D projects originated from a big scale M&A has higher probability of success than from small scale M&A (usually small Biotechs, example of vertical M&A). Furthermore, the results suggest that despite the firms strategically use external R&D to strengthen the R&D portfolio

in a specific ATC1 or diversify to different ATC1, these strategies are not predictors of innovation performance.

4.2 Literature Review

What are the trade-offs among making, allying, and buying? Economic theory suggests that firms ally or buy technology if gains from trade more than compensate transaction costs and asymmetric information problems, such as moral hazard and adverse selection. Ultimately, it is an empirical question if gains from trade dominates or not the loss due transaction costs and asymmetric information problems.

Previous literature [Pisano (1997), Danzon et al. (2005), Guedj (2005), Arora et al. (2009)] compared the innovation performance between making and allying without controlling explicitly for M&A (buying). They don't have a variable that marks the R&D projects acquired via M&A. They compute substantial part of R&D projects originated by M&A in the category in-house projects, because the organizations (Thomson Reuters, Citeline, Adis International, and etc), who compile information about R&D projects, update the datasets in a way that the targeted firm was always part of the acquiring firm. Little attention was given to the innovation performance of R&D projects originated from M&A. The exception was the study of Grabowski and Kyle (2008) that indicates that post-merger projects have higher probability of success than pre-merger projects, controlling only for firm size. It is not clear that result is robust for other controls such as: therapeutic class, innovativeness, risk, chemical/biologic compound, firm and time fixed effects. The problem is that the treatment variable (M&A) cannot be randomized, but it is firms' choice, and it is hard to find a reliable counterfactual to make the comparison between pre and post-merger.

Concerned to the comparison between innovation performance of making and allying, the literature provides mixed evidence. With an old and small sample of 260 biotechnology projects, Pisano (1997) found that collaborative projects have lower success rate than vertically integrated projects. Pisano (1997) argues that Biotechs take advantage of asymmetric information and license-out the "lemons" compounds. In a sample of 4057 R&D projects initiated by 40 big pharmaceutical firms between 1984-2001, Guedj (2005) found that research alliance projects have higher success rate in moving from Phase I to Phase II, but lower success rate in moving from Phase II to Phase III and

in receiving FDA approval. Guedj (2005) estimates that research alliance projects (represents about 10% of the sample) are 10% less likely to receive FDA approval than internal projects. The explanation of Guedj (2005) for this result is that entrepreneurs of Biotechs have personal incentives to continue low quality projects in the early clinical trials, in order to keep raising external funding; whereas managers of internal projects in a Big Pharma take decisions about project continuation based on scientific, clinical and financial criteria. In fact, financial constraints affect the performance of alliances. In a sample of 200 R&D alliances, Lerner et al. (2003) found that alliances signed under periods of limited equity financing are less successful in generating drug approval than alliances signed during favorable financing conditions.

However, in a sample of about 1900 compounds developed by over 900 firms between 1988 to 2000, Danzon et al. (2005) found that in-licensed compounds (represents about 50% of the sample) have higher success rate than in-house compounds. It is worth mentioning that the variable "in-licensed compounds" of Danzon et al. (2005) includes any sort of alliance signed in phase 1, 2, and 3, such as marketing and distribution agreements. According to the Danzon et al. (2005), licensee's experience in drug development increases the success rate, especially phase III, which is more complex and expensive. For Danzon et al. (2005), the gains of trade more than compensate the potential negative effects of moral hazard and adverse selection.

In a sample of over 3000 R&D projects initiated by 329 firms between 1980 and 1994, Arora et al. (2009) found that in-licensed compounds (represents about 8% of the sample) have the same probability of success as in-house compounds, controlling for firm type, scale, scope, economic selection (market size, and market competition), ATC, and diseases characteristics. Arora et al. (2009) added that comparing to big Pharmas, Biotechs have on average lower quality compounds. In this case, lower probability of success of compounds originated by Biotechs can be explained by the average lower quality instead of assuming that Biotechs are on purpose dumping "lemons" compounds to Big Pharmas.

4.3 Hypothesis

Research alliances allow firms to explore complementary technology, expertise, research tools and material, that are not available in-house repertoire. Given the fast pace of technological change,

even for a big firm with stronger research capability like Merck, it might be more cost effective and less risky to learn by collaboration than trying to develop all the capabilities internally [Arora and Gambardella (1990), Gambardella (1992)]. As alliance is a long term targeted effort, the synergy created by a combination of different capabilities and resources tends to overcome the transaction costs and asymmetric information problems. Therefore, the following hypothesis is formulated:

Hypothesis 1a: R&D projects developed by research alliance have higher innovation performance than R&D projects developed internally.

Firms might face lower transaction cost and less asymmetric information problems, if they have stronger experience (measured by previous drug approval) in the same ATC1 of the research alliance. Although the big pharmaceutical firms have internal or external R&D projects in almost all ATC1, they are specialized and have comparative advantage in certain ATC1. For example, based on in-house R&D program, Pfizer has comparative advantage in ATC N (Nervous System); whereas Merck has comparative advantage in ATC A (Alimentary Tract and Metabolism). Probably, a biotech or other firm would have more incentives to dump a "lemon" on Pfizer in ATC A than ATC N.

Hypothesis 1b: Horizontal research alliances have higher innovation performance than vertical research alliances.

Several authors claim that small biotech firms are more innovative than big pharmaceutical firms. A flexible structure with more freedom and less bureaucracy would foster creativity. In fact, small firms have more patents and publication normalized by total assets. However, in the pharmaceutical industry, it is necessary to distinguish between invention and innovation. From invention (patents) to innovation (bring the drug to the market), there is a long shot, that is, the compound has to pass the clinical trials, showing safety and efficacy, and this process can last several years. Other authors [Pisano (1997), Guedj (2005)] claim that biotech firms keep the higher quality compounds and dump the lower quality compounds on big pharmaceutical firms. Arora et al. (2009) found that big pharmaceutical firms are more innovative than small biotech firms, because on average the first has higher quality compounds than the second. Furthermore, a partnership between two big firms might be more successful because of the higher probability of combining complementary

capabilities and resources, and both parties might have less incentives to "free ride", in order to preserve the reputation. Therefore, the following hypothesis states that the partner size affects the innovation performance of a research alliance.

Hypothesis 1c: Research alliances with a big firm have higher innovation performance than research alliances with a small-medium firm.

Quite often, M&A is described in the literature as short term solution for a more structural problem related to the R&D pipeline crisis [Higgins and Rodriguez (2006)]. For antitrust purpose, firms often claim that M&A creates economies of scale and scope at R&D level, but some authors suggest [Mittra (2007)] a diseconomies in basic research and potentially economies at marketing and distribution level. M&A might also have other rationality than buying R&D projects or technology not available in-house. Firms might use M&A to promote organizational change, in order to reduce operating cost and to shut down inefficient units or operations. According with a case study of Mittra (2006), firms might also engage in M&A to access global markets (Ciba-Geigy and Sandoz in 1996, Rhone-Poulenc Rorer and HoechstAG in 1999; and Sanofi and Aventis in 2004). The literature didn't pay much attention, but M&A can be employed to obtain tax advantage, to write-off extremely large expenses, and to change the tax domicile. Therefore, the following hypothesis is formulated:

Hypothesis 2a: R&D projects acquired via M&A have lower innovation performance than R&D projects developed internally.

Given a specific M&A, some acquired R&D projects might belong to the ATC1, in which the firm has stronger experience (previously drug approval) and others not. Based on the same logic of horizontal and vertical research alliances, R&D projects that explore higher level of complementarity and strengthen the ATC1, in which the firm is specialized, are more likely to be successful. Instead of relying on organic growth, firms might diversify and acquire R&D projects in ATC1 outside their core competence [Arora and Gambardella (1990)]. In fact, there is also complementarity or economies of scope between different ATC. It is an empirical question what M&A has higher innovation performance, vertical or horizontal. However, transaction costs and asymmetric information might be more problematic in the case of vertical M&A, in which firms are acquiring R&D projects outside of their core competence. Therefore, the following hypothesis is formulated:

Hypothesis 2b: R&D projects originated from horizontal M&A have higher innovation performance than vertical M&A.

Based on the same rationality of hypothesis 1c, R&D projects originated from a big firm, that is, a big M&A (more than 100 R&D projects and the deal over U$ 1 Billion) might have on average higher probability of success than R&D projects originated from small firms (Biotechs). Several factors might explain this difference in innovation performance, but it looks that R&D portfolio of big firms is on average higher quality than small firms. It is worth mentioning that if small M&A is often related to technology acquisition, big M&A is usually driven by a set of different rationality as described previously. In this line of reasoning, R&D projects originated from small M&A should have higher probability of success than R&D projects originated from big M&A, but it might be the case that several big firms use M&A to buy specific research tools or material to be incorporated to their own R&D projects instead of developing the R&D projects of Biotechs. Therefore, the following hypothesis is formulated:

Hypothesis 2c: R&D projects acquired via big M&A have higher innovation performance than small M&A.

4.4 Data

4.4.1 Data Source and Sample

The main data source is the Pharmaprojects database from Citeline, an organization that claims to be the world's most comprehensive source of real-time R&D intelligence for the pharmaceutical industry. Pharmaprojects tracks drugs pipeline since 1980, covering all stages of drug development from discovery through clinical trials to drug approval and commercialization. Besides providing a detailed development history timeline of over 60,000 drug profiles, Pharmaprojects includes licensing and sponsoring organization details together with a description of the disease, therapy, mechanism, delivery routes, chemical data and molecular structure.

To keep the sample homogeneous and focus on firms that engage in research alliance and M&A, the sample is restricted to 32 firms with highest number of R&D projects among 1980 to 2014. Only new chemical entity (NCE) and biologic compounds are included in the sample. R&D projects

without a well-defined chemical or biological molecular structure were eliminated, such as natural product originated by animals and plants. The sample is also restricted to R&D projects with entering date from 1989 to 2007, in order to mitigate the left and right censoring problems. For example, left censoring occurs when drug approval is reported but with no information about clinical trials or the project starting date. Right censoring happens in the case of no information about project success or failure, although it is known that the R&D project started or passed some clinical trials. Furthermore, the lower bound 1989 is chosen because Citeline started a more systematic and standardized data collection after 1988 and most part of observations occurred after this date. The upper bound is chosen because the development of R&D projects take long time, and if R&D projects started in 2008 to 2014 were included, the rate of failure would be biased upward.

Other datasets used for this study are LexisNexis® Academic, the Pharmaceutical Market from OECD (The Organisation for Economic Co-operation and Development), and MalaCards from Weizmann Institute of Science in Israel. The first is an online database with legal and business information from over 17,000 sources that it is used to generate a list of M&A. The second database is used to construct the variable Market Size for 9 ATC (Anatomical Therapeutic Chemical Classification System) level 1: A (Alimentary tract and metabolism), B (Blood and blood forming organs), C (Cardiovascular system, G (Genito-urinary system and sex hormones), H (Systemic hormonal preparations), J (Antiinfectives), M (Musculo-skeletal system), N (Nervous system), and R (Respiratory system). The third is a searchable database of over 18,864 human disease entries from 64 consolidated sources that it is used to classify diseases from Pharmaproject in categories such as: Cancer, Rare (or Orphan Disease), Genetic, and Infectious.

The final sample has 14,861 observations, 9,201 R&D projects, 710 different indications/diseases in the most disaggregate level, 14 ATC level 1 (group of organ/system like M - Musculo-skeletal system), and 147 ATC level 2 (indication like M05 - Drugs for treatment of bone diseases, ex: Osteoporosis). Although the most part of variation is at R&D project level given how the data was collected and overwritten by Citeline, the unit of analysis is a specific indication as Danzon et al. (2005), Arora et al. (2009). For example, the R&D project or compound Sildenafil (Trademark: Viagra) developed by Pfizer targeted four indications: 1) male erectile dysfunction; 2) pulmonary arterial hypertension; 3) female sexual dysfunction; and 4) surgery adjunct. Each indication had to pass a separate clinical trial and in this case only indications 1 and 2 were successful in receiving

Table 4.1: Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
Approval	14861	0.046	0.21	0	1
External	14861	0.59	0.49	0	1
Research Alliance	14861	0.14	0.35	0	1
M&A	14861	0.53	0.50	0	1
Vertical Alliance	14861	0.060	0.24	0	1
Horizontal Alliance	14861	0.082	0.27	0	1
Alliance_Small_Firm	14861	0.092	0.29	0	1
Alliance_Big_Firm	14861	0.051	0.22	0	1
Vertical M&A	14861	0.19	0.39	0	1
Horizontal M&A	14861	0.27	0.44	0	1
Small M&A	14861	0.15	0.36	0	1
Big M&A	14861	0.30	0.46	0	1
Biologic	14861	0.19	0.39	0	1
Innovation Index	14861	0.30	0.31	0.053	1
# of Diseases by Project	14861	3.10	3.59	1	29
University Project	14861	0.019	0.14	0	1
Vaccine	14861	0.042	0.20	0	1
Scale Firm	14861	3.67	0.86	0	5.00
Scope Firm	14861	10.6	3.13	2	14
Risk Portfolio	14861	0.93	0.039	0.39	1
Screening Ratio	14861	0.43	0.21	0	1
Cancer	14861	0.20	0.40	0	1
Rare	14861	0.046	0.21	0	1
Genetic	14861	0.16	0.37	0	1
Infectious	14861	0.14	0.35	0	1
Competition	14861	23.3	24.3	0	119
Market Size	12010	12.5	6.29	1.94	32.1

approval from regulatory authority. Therefore, each indication is considered as a separate R&D project.

4.4.2 Summary Statistics

Summary statistics are presented in Table 4.1. The dependent variable "Approval" is coded 1 if the indication was approved by any regulatory authority in the world, and 0 otherwise. About 4.6% of indications was successful in receiving approval from regulatory authority.

The variable of interest "External" is coded 1 if the R&D project was originated/developed by research alliance or M&A, and 0 if the R&D project was completely developed in-house. External R&D and Internal R&D represent respectively 60% and 40% of total R&D projects. Each R&D

project was classified as research alliance or M&A based on project description (field "Overview") and several other fields from Pharmaprojects, and a list of M&A from LexisNexis® Academic. For example, the blockbuster drug Enbrel indicated for arthritis, ankylosing spondylitis, and psoriasis was coded as research alliance not only because the field "Originator" and "Licensee" had respectively the names "Amgen" and "Pfizer/Takeda/Sanofi", but the project description used the words "jointly developed":

> "*Etanercept (Enbrel) is a soluble recombinant human p75 tumour necrosis factor (TNF) receptor and human IgG1 Fc portion fusion protein produced in a mammalian cell expression system, jointly developed by Receptech (Amgen; previously Immunex) and Wyeth Pharmaceuticals (Wyeth (Pfizer) (formerly Wyeth-Ayerst (American Home Products (AHP)) for the treatment of rheumatoid arthritis (RA) and other inflammatory conditions ...*"

Quite often, there were many firms in the "Licensee" field, but no reference in the project description field for words like "co-developed", "joint venture", joint developed, and etc. In this case, "Licensee" is less likely to be a research alliance and more likely to be a marketing, promotion, or distribution agreement, what can be confirmed by descriptions in the fields "Marketing" and "Licensing" from Pharmaprojects. In the project description field, it is also possible to track M&A. For example, Receptech was a subsidiary of Immunex, in which was acquired by Amgen in 2002; and American Home Products was renamed to Wyeth in 2002 and was acquired by Pfizer in 2009.

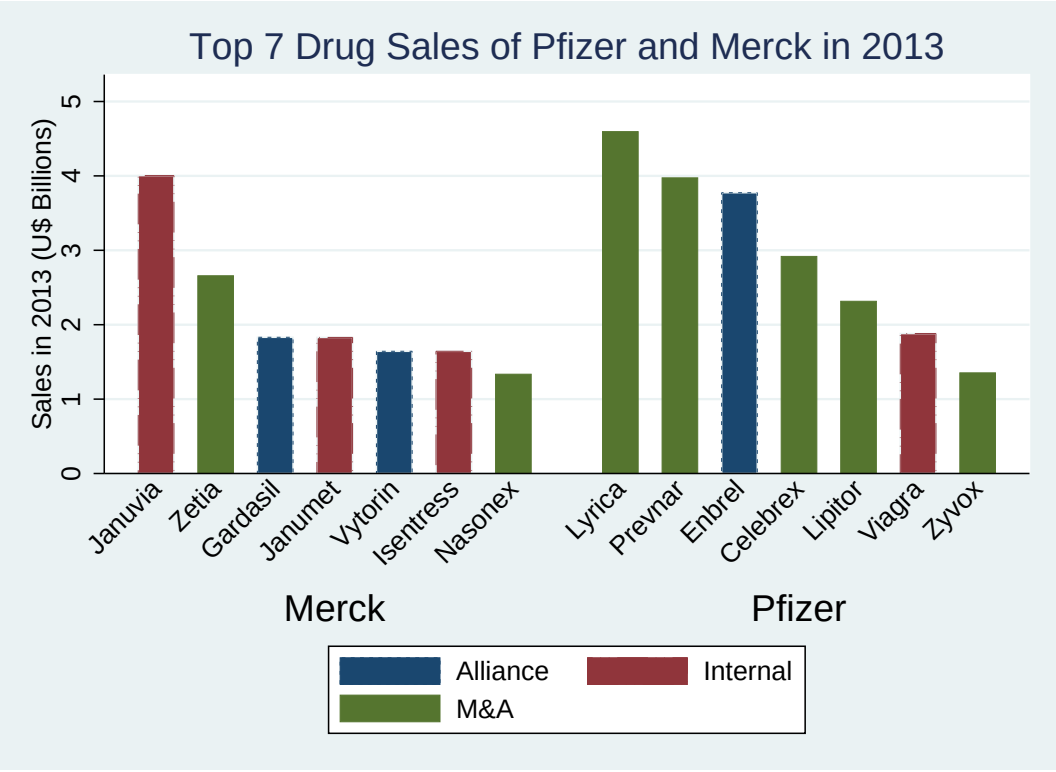
The variable "Research Alliance" indicates that 14% of the R&D projects were developed in research alliance, a percentage slightly higher than described by Guedj (2005), Arora et al. (2009), but substantial lower than 50% informed by Danzon et al. (2005). Following Guedj (2005), Arora et al. (2009), this study focus on research alliance; whereas Danzon et al. (2005) included marketing, promotion, and distribution agreement under the variable "Alliance". The variable "Research Alliance" was split in other four variables "Vertical Alliance" vs "Horizontal Alliance", and "Alliance_Small_Firm" vs "Alliance_Big_Firm". The variable "Vertical Alliance" indicates that the research alliance was formed in a ATC1, in which a firm has no drug approval originated by internal R&D; whereas the variable "Horizontal Alliance" indicates that the research alliance was formed in an ATC1, in which a firm has already a drug approval originated by internal R&D. The variable "Alliance_Big_Firm" was coded as 1, if the licensee was one of the 32 biggest firms in terms or R&D projects, and 0 otherwise. The variable "Alliance_Small_Firm" was coded as 1, if

the licensee wasn't the one of the 32 biggest firms in terms or R&D projects, and 0 otherwise.

R&D projects originated from M&A represent 53% of the sample. Although the specialized literature [Danzon et al. (2005), Guedj (2005), Arora et al. (2009)] doesn't control explicitly for M&A, Danzon et al. (2005) recognizes that the omission might introduce measurement error in the variable of interest. The variable "M&A" was also split in four variables "Vertical M&A" vs "Horizontal M&A", and "Small M&A" vs "Big M&A". The variable "Vertical M&A" indicates that the firm acquired R&D project in the ATC1, in which the firm had no drug approval originated by internal R&D; whereas the variable "Horizontal M&A" indicates that the firm acquired R&D project in the ATC1, in which the firm had at least one drug approval originated by internal R&D. The variables "Small M&A" and "Big M&A" were created based on the size of M&A, that is, big M&A was defined as the acquisition of more than one hundred R&D projects and the value of targeted firm above U$ 1 billion; whereas the small M&A was the opposite.

Fig 1 presents the Top 7 Drug Sales of Pfizer and Merck in 2013 by R&D Project Source. Among the big pharmaceutical firms, Merck has the reputation of having stronger internal R&D capabilities; whereas Pfizer has the reputation of relying stronger on external R&D. According with Fig 1, even Merck relays considerable on external R&D. Among the top 7 drug sales in 2013, Merck developed 3 drugs internally: Januvia, Janumet, and Isentress. Zetia and Nasonex were developed by Schering-Plough, a firm acquired by Merck in 2009. Gardasil was developed in collaboration with the University of Queensland in Australia. Vytorin was developed in joint-venture with Schering-Plough before the acquisition of Schering-Plough by Merck. Among the top 7 drug sales in 2013, Pfizer only developed internally the Viagra. Lyrica was developed by Parke-Davis, a firm acquired by Warner–Lambert in 1970, which in turn was acquired by Pfizer in 2000. Prevnar was developed by Wyeth, a firm acquired by Pfizer in 2000. Celebrex was developed by Searle, a firm acquired by Monsanto in 1985 and merged with Pharmacia & Upjohn in 1999, which in turn was acquired by Pfizer in 2003. Lipitor was developed by Warner–Lambert, a firm acquired by Pfizer in 2000. Zyvox was developed by Pharmacia, a firm merged with Upjohn in 1999 and acquired by Pfizer in 2003. Enbrel was jointly developed by Immunex and Wyeth. Immunex was acquired by Amgen in 2002; whereas Wyeth was acquired by Pfizer in 2000.

The control variables at project level are: Biologic, Innovation Index, # of Diseases by Project, University Project, and Vaccine. The variable "Biologic" is coded 1 if the compound is produced



Source: Merck (2014), Pfizer (2014)

Figure 4.1: Top 7 Drug Sales of Pfizer and Merck in 2013 by R&D Project Source

by living organism (such as yeast, bacteria, and etc) and not made by chemical synthesis (small molecules), and 0 otherwise. About 20% of compounds is biological and this proportion has been increasing during the years. The variable "Innovation Index" was constructed based on Guedj (2005) and aims to ranking R&D projects in terms of chronological use of drug's mechanism of action in the body. For example, the pharmacological mechanism of Enbrel is "tumour necrosis factor alpha antagonist". This mechanism of action was already used in another R&D project started in 1989, in which received rank 1. As Enbrel project started in 1990, it received rank 2. R&D projects that used the same mechanism and started in 1991 and 1992 received respectively rank 3 and 4, and so on. The variable "Innovation Index" is the inverse of the chronological rank, therefore the maximum value 1 represents that the firms used the most state of art technology/knowledge available when the project started, and the minimum value in the sample .053 indicates that the project used an old technique/knowledge described in the pharmacological literature. The variable "# of Diseases by Project" might captures different factors such as spillover between therapeutic class, project size, and unobservable quality of the project. For example, Enbrel project targets 18 indications in 9 different therapeutic class. The variable "University Project" is coded 1 if a university or non-profit research institute formally participate in R&D project, and 0 otherwise. Only 2% of R&D projects has a participation of a university or non-profit research institute. However, the literature [Ward and Dranove (1995), McMillan et al. (2000), Zucker et al. (2002), Stuart et al. (2007)] points out for the importance of the upstream inputs (patents, human capital, star scientist, research tools, and etc) as source of competitive advantage in the discovery phase of drug development. The variable "Vaccine" is coded 1 if the project is developing a vaccine, and 0 otherwise. Only 4.2% of projects are developing a vaccine.

The control variables at firm level are: Scale Firm, Scope Firm, Risk Portfolio, and Screening Ratio. The variable "Scale Firm" is the total number of R&D projects per year by the firm. The variable "Scope Firm" is the number of ATC1, in which the firm obtained at least one drug approval from 1989 to 2014. Firms such as Kyowa Hakko Kirin, Servier, and Shire, only obtained drug approval in two different ATC1; whereas only two firms Novartis and Sanofi obtained drug approval in all 14 ATC1. The variable "Risk Portfolio" is the mean of diseases attrition rate per year by firm. Each disease has a historical attrition rate, for some diseases there are several approved drugs, but for others zero. The maximum value 1 for this variable means that in certain year a firm started projects

targeting diseases without any drug approval. The minimum value .389 means that a firm carried a low risk portfolio targeting diseases that is relatively easy to receive drug approval. The mean of this variable indicates that typical portfolio carries projects with 92% probability of fail. The variable "Screening Ratio" follows Danzon et al. (2005), and it is the percentage of projects that a firm takes into clinical trial by year. A lower screening ratio means that a firm might be very selective and discontinue projects before the clinical trial. By the other hand, higher screening ratio indicates that firm might advance even project with poor prospect to the clinical trial. Alternative explanation is that lower screening ratio indicates poor quality R&D portfolio; whereas higher screening ratio indicates that the firm has stronger innovation capabilities with several higher quality R&D projects in the discovery phase. About 43% of all projects starts phase 1 of clinical trial, that is, the most part of R&D projects is terminated before the compound being test in human beings.

The control variables at diseases level are: Cancer, Rare, Genetic and Infectious. They are dummy variables and non-mutually exclusive categories. Besides controlling for ATC1 and ATC2, this study control for a group of diseases that supersede the taxonomy based on anatomy, such as cancer that can be in any therapeutic class.

For robustness checks, two more variables are used: Competition and Market Size. The variable "Competition" is the logarithm of the total number of drug approval +1 during 1980 to 2014 for a specific condition (diseases). The variable "Market Size" is the logarithm of drug sales by ATC1 in six countries without missing values between 1989-2007: Finland, France, Germany, Japan, Netherlands, and Switzerland. All sales data was in U$ dollars (Billions) and was adjusted for inflation base year 2007, using the CPI (Consumer Price Index) provided by US Bureau of Labor Statistics. Unfortunately, OECD doesn't have data for United States, the biggest market. Dubois et al. (2015) point out that both variables are related and might affect the innovation rate in a non-linear way or the causality might be in both direction. A large expected market size creates more incentives for innovation, but innovation also creates market size. Competition decreases the profits, what can decrease the incentives to innovate, but firms might respond to competition with more innovation, in order to create a new market and profits. A large market size also induces entry (more competition), then a disease with a large market size with many competitors might be less profitable than an orphan disease with a lower market size and few competitors.

Table 4.2 reports the probability of drug approval for each R&D source. Internal and External

Table 4.2: Probability of Drug Approval

	Obs	Mean
Internal	6047	0.042
External	8814	0.049
Research Alliance	2112	0.074
M&A	6702	0.041
Vertical Alliance	895	0.080
Horizontal Alliance	1217	0.069
Alliance_Small_Firm	1361	0.036
Alliance_Big_Firm	751	0.14
Vertical M&A	2762	0.036
Horizontal M&A	3940	0.044
Small-Medium M&A	2255	0.028
Big M&A	4447	0.047

source have slightly different probability of success respectively 4.2% and 4.9%, and this difference is not statistically significant. However, when the variable "External" is split in different categories, higher variation in the probability of success rate appears. The probability of drug approval for research alliance and M&A is respectively 7.4% and 4.1%. When the variable "Research Alliance" is split in "Vertical Alliance" and "Horizontal Alliance", the difference (1.1%) between probability of success is not very high, but if the variable "Research Alliance" is split in "Alliance_Small_Firm" and "Alliance_Big_Firm", the difference (10.4%) between probability of success is very significative. About 1/3 of research alliances is between big firms and the probability of drug approval of these alliances is 14%; whereas research alliances with small firms (licensee is not one of 32 biggest firms in terms of R&D projects) have probability of drug approval of 3.6%. Similar to the variable "Research Alliance", when the variable "M&A" is split in "Vertical M&A" and "Horizontal M&A", the difference (0.8%) between probability of success is not very high. If the variable "M&A" is split in "Small M&A" and "Big M&A", the difference (1.9%) between probability of success is not very substantial either. However, the pattern of splitting the variables "Research Alliance" and "M&A" suggests that the size of the firm matters more than the strategic decision about specialization vs diversification for the success rate of drug approval.

Next section will show that these differences in the probability of drug approval remain even after controlling for a set of control variables described previously in this section.

Table 4.3 reports the correlation between variables. There is no evidence of multicollinearity

problem. Overall, the sign of coefficients match with the results in the next section.

4.5 Empirical Strategy

Equation 4.1 represents the basic specification to test the difference of innovation productivity among R&D projects originated/developed by internal R&D, alliance, and M&A.

$$Approval_{pfay} = \beta_0 + \beta_1 Alliance_{pfay} + \beta_2 M\&A_{pfay} + \theta_i X_{pfay} + \phi_f + \alpha_a + \gamma_y + \varepsilon_{pfay} \quad (4.1)$$

The dependent variable, $Approval_{pfay}$, is a dummy variable that captures the success/failure of the project p, by firm f, in anatomical therapeutic chemical class a, initiate in year y. This dummy variable has value of 1 if the drug was approved by any regulatory authority in the world, and 0 otherwise.

The independent variables of interest are: $Alliance_{pfay}$ and $M\&A_{pfay}$. The category base is internal project. Therefore, the project p could be originated/developed by internal R&D, or external R&D (research alliance or M&A). For example, $Alliance_{pfay}$ and $M\&A_{pfay}$ are respectively dummy variables that have value 1 if the project was developed/originated in alliance/M&A with/from another firm, and get the value 0 otherwise.

The vector of control variables is represented by X_{pfay}, in which $\theta_i s$ are the different coefficients for each control variable. A set of control variables at project level p and firm level f is tested and described in details in the summary statistics. The logic is to show that after controlling for several important factors described in the literature, the variation in innovation productivity can only be explained by R&D sources.

As usual, ϕ_f are dummies for firms, α_a are dummies for Anatomical Therapeutic Chemical (ATC), γ_y are dummies for years, and ε_{pfay} is the error term. Part of variation of innovation productivity is firm specific, for example how they organize R&D program, combine internal R&D with external R&D, explore spillovers from basic science, and design incentives to the scientists. Specifically, firms self-select R&D strategies, in which they can create more value. Merck is known to have stronger internal R&D capabilities; whereas Pfizer rely more on research alliances and M&A.

Table 4.3: Correlation Matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Approval	(1)	1.00												
External	(2)	0.02	1.00											
Research Alliance	(3)	0.05**	0.34**	1.00										
M&A	(4)	-0.01	0.88**	0.01	1.00									
Biologic	(5)	0.08**	0.22**	0.15**	0.20**	1.00								
Innovation Index	(6)	0.05**	0.04**	-0.00	0.05**	0.02**	1.00							
# of Diseases by Project	(7)	0.13**	0.05**	0.16**	-0.03**	0.02*	0.08**	1.00						
University Project	(8)	-0.00	0.01	0.00	0.01	0.06**	0.05**	-0.01	1.00					
Vaccine	(9)	0.10**	0.10**	0.02*	0.10**	0.42**	-0.09**	-0.06**	0.05**	1.00				
Scale Firm	(10)	-0.00	0.11**	-0.08**	0.15**	-0.06**	0.02*	0.06**	-0.03**	0.02*	1.00			
Scope Firm	(11)	0.00	0.07**	-0.05**	0.10**	-0.00	0.04**	0.01	-0.03**	0.05**	0.76**	1.00		
Risk Portfolio	(12)	-0.17**	-0.04**	-0.02	-0.03**	-0.10**	0.06**	-0.05**	-0.00	-0.21**	0.04**	0.02*	1.00	
Screening Ratio	(13)	0.06**	0.03**	0.15**	-0.04**	0.12**	-0.11**	0.24**	-0.02	0.06**	-0.06**	-0.03**	-0.36**	1.00

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

As unobserved factors to the econometrician affect firm's decision among making, allying, and buying, the impact of the variables $Alliance_{pfay}$ and $M\&A_{pfay}$ might be interpreted conditioned firms decide to engage in alliances and M&A. However, these unobserved factors are likely to be firm specific and can be controlled by firm dummies. The idea is to show that results are not driven by a specific firm or heterogeneity among firms.

Different ATCs also present different drug approval rates. ATC classes vary in terms of profitability, market size, competition, opportunity technology, and etc. Therefore, it is important to control for ATC class, mitigating the possibility that some unobserved factor biases the results. As robustness checks, this study will control for higher ATC level and the most possible disaggregate level, that is, disease level.

Finally, a set of time dummies is used to control for external shocks, some disruptive technological change that could affect the innovation productivity, R&D cycle, or trends, for example, the literature point outs a "productivity crisis" in the past 30 years, in which firms have been spending more and more in R&D each year, but the number of drug approval has been decreasing [DiMasi et al. (2003), DiMasi and Grabowski (2007)].

Someone might argue that there is a reverse causality between the dependent variable, $Approval_{pfay}$, and the independent variable, $Alliance_{pfay}$. This potential source of endogeneity is unlikely to be a problem in this study, because of the long lag between research alliance and drug approval. Research alliance usually happens in the discovery phase, several years before the drug approval. The future expectation of drug approval may affect the alliance formation. However, we control for several factors related to the expectation of drug approval, such as ATC classes, diseases, market size, and risk portfolio.

Equation 4.1 can be estimated by assuming nonlinear model (discrete choice model as Probit) and OLS (Linear Probability Model). It is assumed that unobserved factors are not correlated with the variables of interest. As robustness checks, several variations of equation 4.1 are estimated in the result section. Robust standard errors are clustered by firm level.

We also split the variable: 1) $Alliance_{pfay}$ in "Vertical Alliance" vs "Horizontal Alliance", and in "Alliance_Small_Firm" vs "Alliance_Big_Firm"; 2) $M\&A_{pfay}$ in "Vertical M&A" vs "Horizontal M&A", and in "Small M&A" vs "Big M&A". We estimated the relative innovation performance of the 4 groups using the Propensity Score Matching (PSM) method [Rosenbaum and Rubin (1983),

Dehejia and Wahba (2002), Cameron and Trivedi (2005), Abadie and Imbens (2016)]. PSM is an inexact matching estimator designed to deal with observational data. The matching is not made directly on the vector of control variables X_{pfay}, but made on the propensity score, $p(x)$, the conditional probability of receiving theatment given the control variables X_{pfay}. As the vector of control variables X_{pfay} has a high dimension, and is composed by several continuous variables, the PSM is the most appropriate matching technique for this study. However, other matching techniques like Nearest-Neighbor Matching estimator (NNME) don't change the results. The advantage of PSM over OLS is the less restrictive functional form for identification, but it requires bigger sample size.

To identify the population-average treatment effect on the treated (ATET), it is necessary to assume the overlap assumption. For example, for each propensity score of "Vertical Alliance", there is another similar propensity score of "Horizontal Alliance". Propensity scores were generated using a Probit model. Concerned to the number of matches per observation, there is a tradeoff between bias and variance [Cameron and Trivedi (2005)]. Higher number of matches per observation decreases the variance, but it also increases the probability of inferior matches. To minimize the bias, only one match per observation was allowed. The robust standard errors were derived by Abadie and Imbens (2006, 2012).

4.6 Results

Table 4.4 reports the probit estimates for the probability of drug approval. Robust standard errors are clustered by firms for all regressions. Firm, disease group (cancer, rare, genetic and infectious), ATC1, and year dummies are presented in all regressions. Column 1 shows that the probability of drug approval of external R&D projects is not statistically different compared to internal R&D projects, the reference category. Column 2 shows that even controlling for a set of variables at project and firm level, the result of column 1 remains. Both results are aligned with Arora et al. (2009) who claim that internal projects have the same probability of success as licensed projects. However, when the variable of interest "External" is split in "Research Alliance" and "M&A" in columns 3 to 5, a consistent pattern indicates that research alliances have higher probability of drug approval than internal R&D projects; whereas R&D projects acquired by M&A have lower probability of success than internal R&D projects. The results of Table 4.4 also contradict Pisano (1997),

Guedj (2005) who claim that internal projects have higher probability of success than projects developed in research alliance. One explanation for the results obtained by Arora et al. (2009), Pisano (1997), Guedj (2005) is that these studies don't control explicitly for M&A, therefore their estimations of the impact of research alliance are likely to be downward biased as suggested by the statistically significant negative coefficient of "M&A" in regressions 3 to 5.

Concerned the control variables at R&D project level, regressions 2, 4, and 5 show that biologic compounds have higher probability to receive drug approval than small molecules (chemical compounds). Some authors suggest that biologic compounds embody higher level of innovation, even though the cost and time for development is similar to the chemical compounds [DiMasi et al. (2003), DiMasi and Grabowski (2007), Trusheim et al. (2010)]. The coefficient of "Innovation Index" is not statistically significant. As the sample is restricted to the new chemical entity (NCE) and biologic compounds, there is no much variation in terms of innovativeness captured by the variable "Innovation Index". The coefficient of "# of Diseases by Project" is positive and statistically significant. This variable is likely to captures size, spillovers between ATC1, and unobservable quality of R&D projects. Very similar variable was created, the number of ATC1 by project, and the results remain the same. The coefficient of "University Project" is negative and statistically significant, but it is not statistically significant in the robustness checks of Table 4.5. Several authors [Cockburn and Henderson (1998), Zucker et al. (2002), Stuart et al. (2007)] claim that proximity to basic science measured by star scientists, publication, and partnership with universities foster innovation. There is some empirical evidence that spillovers from basic science matter for invention such as number of patents, but in the pharmaceutical industry there is a long shot between invention and innovation. Some studies [Schulze and Ringel (2013), Roland et al. (2015)] state that first to patent is not a predictor to get drug approval or the best drug in class. The coefficient "Vaccine" is not statistically significant, but it is worth mentioning that 98% of vaccine in the sample is classified as biologic compound.

Related to control variables at firm level, regressions 2 and 5 indicate that economies of scale doesn't explain differences in innovation productivity. However, this variable is important in studies [Danzon et al. (2005), Arora et al. (2009)] that have hundreds of small firms (biotechs) in the sample. As the sample of this study is composed only by big pharmaceutical firms, it might be the case that the firms are operating in an optimal scale or that size matters only at project level

Table 4.4: Baseline Estimation for the Probability of Drug Approval

	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) Probit
External	0.02	-0.06			
	(0.08)	(0.07)			
Research Alliance			0.31***	0.24***	0.26***
			(0.07)	(0.07)	(0.07)
M&A			-0.15*	-0.22**	-0.22**
			(0.08)	(0.09)	(0.09)
Biologic		0.31***		0.36***	0.33***
		(0.07)		(0.07)	(0.07)
Innovation Index		0.08		0.04	0.08
		(0.11)		(0.12)	(0.12)
# of Diseases by Project		0.11***		0.11***	0.10***
		(0.02)		(0.02)	(0.02)
University Project		-0.24*		-0.31**	-0.27*
		(0.14)		(0.14)	(0.14)
Vaccine		0.19		0.26	0.20
		(0.22)		(0.22)	(0.23)
Scale Firm		-0.09			-0.08
		(0.08)			(0.08)
Scope Firm		-0.17***			-0.17***
		(0.03)			(0.03)
Risk Portfolio		-3.18***			-3.45***
		(0.75)			(0.72)
Screening Ratio		0.31			0.25
		(0.22)			(0.22)
Firm Dummies	Yes	Yes	Yes	Yes	Yes
Disease Group	Yes	Yes	Yes	Yes	Yes
ATC1 (14 Dummies)	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Observations	14861	14861	14861	14861	14861

Note: Robust standard errors are in parentheses (clustered by firms).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.5: Robustness Checks for the Probability of Drug Approval

	(1) Probit[1]	(2) Probit	(3) LPM	(4) LPM	(5) LPM
Research Alliance	0.270***	0.319***	0.022***	0.037***	0.015*
	(0.08)	(0.08)	(0.01)	(0.01)	(0.01)
M&A	-0.313***	-0.223**	-0.016**	-0.019**	-0.012*
	(0.08)	(0.10)	(0.01)	(0.01)	(0.01)
Biologic	0.298***	0.398***	0.014*	0.020*	0.011*
	(0.09)	(0.08)	(0.01)	(0.01)	(0.01)
Innovation Index	0.007	0.179	0.001	0.010	0.003
	(0.12)	(0.14)	(0.01)	(0.01)	(0.01)
# of Diseases by Project	0.093***	0.123***	0.009***	0.019***	0.007***
	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)
University Project	-0.123	-0.287*	-0.022	-0.030*	-0.017
	(0.14)	(0.15)	(0.01)	(0.02)	(0.01)
Vaccine	-0.411*	0.108	0.027**	0.029	-0.030*
	(0.23)	(0.24)	(0.01)	(0.03)	(0.02)
Scale Firm	-0.000	-0.089	-0.004	-0.005	-0.007
	(0.10)	(0.08)	(0.01)	(0.01)	(0.01)
Scope Firm	-0.258***	-0.172***	-0.016***	-0.020***	-0.011***
	(0.04)	(0.04)	(0.00)	(0.00)	(0.00)
Risk Portfolio	-3.074***	-3.194***	-0.700***	-0.702***	-0.155
	(1.16)	(0.78)	(0.13)	(0.15)	(0.10)
Screening Ratio	-0.091	0.226	-0.005	-0.006	0.011
	(0.26)	(0.22)	(0.02)	(0.02)	(0.02)
Market Size		-0.036***		-0.002**	
		(0.01)		(0.00)	
Competition		0.001		0.000	
		(0.00)		(0.00)	
Firm Dummies	Yes	Yes	Yes	Yes	Yes
Disease Group	Yes	Yes	Yes	Yes	Yes
ATC1 (14 Dummies)	Yes	Yes	No	No	No
ATC2 (147 Dummies)	No	No	Yes	Yes	No
Diseases (710 Dummies)	No	No	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Observations	9209	12010	14861	12010	14861

Note: Robust standard errors are in parentheses (clustered by firms).

[1] Model 1 - Sample restricted to the R&D projects that at least achieve the preclinical phase.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

as suggested by the positive statistically significant coefficient of "# of Diseases by Project". An increase of "Scope Firm" decreases the probability of drug approval, suggesting that specialization in few therapeutic classes might be a superior strategy than diversification, or that firms are operating above the optimum scope level. Higher "Risk Portfolio" is associated with lower probability of drug approval. This variable controls for risk loving firms that focus on ATC1 or diseases that is difficult to innovate. These firms might prefer to swing for the fences (blockbuster strategy) than hit singles [Arora et al. (2009)]. The coefficient of "Screening Ratio" is positive but not statistically significant, consequently it is not possible to conclude that more compounds brought to clinical trial increase the innovation productivity.

Table 4.5 presents the robustness checks. As the variables of interest ("Research Alliance" and "M&A") are formed by firm choice, there is a potential endogeneity bias, if unobserved factors affect the firm choice and the probability of drug approval. Different specifications, sample restriction, and econometric modelling assumptions such as linear probability model with dummies variables for specific diseases are used to mitigate this potential endogeneity problem. The results of columns 1 to 5 show that the coefficient of "Research Alliance" and "M&A" are statistically significant and has the same sign of regression 5 - Table 4.4, the preferred model.

Some authors point out [Danzon et al. (2005), Arora et al. (2009)] that firms are more likely to kill or advance projects based on economic considerations (market size, complementary downstream assets, competition, product cannibalization, and etc) in the discovery phase. After the discovery phase, technical and scientific considerations would play a more important role than economic incentives. The selection issue (killing or advancing projects based on unobserved factors to the econometrician) is more problematic when the sample is composed by very heterogeneous firms. For example, compared to pharmaceutical firms, biotechs under financial constraints might advance poor quality projects to the clinical trials in order to keep receiving external funding [Guedj (2005)]. In column 1 of Table 4.5, the same regression 5 of Table 4.4 is presented, but with the sample restricted to the R&D projects that at least achieve the preclinical phase, that is, R&D projects discontinued in the discovery phase were excluded. This result has to be interpreted as the probability of success conditioned to pass the discovery phase. Together with the fact that the coefficient of "Screening Ratio" is not statistically significant in any regression of Table 4.4 and 4.5, the selection issue doesn't look to be a problem in this paper.

Regressions 2 and 4 add the variables “Competition” and “Market Size”. Both coefficients are not stable and not always statistically significant in other specifications not showed in this paper. Therefore, no interpretation was provided about how these two variables would affect the probability of drug approval, but the coefficients of “Research Alliance” and “M&A” are still statistically significant and follows the same pattern of Table 4.4. The difference between regressions 2 and 4 is that the first one is a Probit controlling for ATC1; whereas the last one is a Linear Probability Model that accommodates 147 dummies for ATC2.

Regressions 3 and 5 of Table 4.5 are Linear Probability Model that control respectively for ATC2 and 710 dummies for different diseases. The idea is to show that the results are not driven by specific ATC2 or diseases, as the technology opportunity, risk and cost development, profitability varies a lot among diseases. The coefficients of “Research Alliance” and “M&A” of Regression 5 are less statistically significant and with lower magnitude than of Regression 3, what makes sense as the last regression control for more specific diseases level, removing part of variation. The most conservative estimation (regression 5) indicates that: 1) research alliance has probability of drug approval 1.5% higher than internal R&D projects; and 2) R&D projects originated from M&A has probability of drug approval 1.2% lower than internal R&D projects.

Table 4.6 reports the probit estimates for the probability of drug approval. Column 1 split the variable “Research Alliance” in “Vertical Alliance” and “Horizontal Alliance”. Both coefficients are statistically significant and positive with similar magnitude. However, when the variable “Research Alliance” is split in “Alliance_Big_Firm” and “Alliance_Small_Firm” in column 2, the first coefficient is statistically significant and positive with a comparative higher value; whereas the second coefficient is negative and statistically significant. These results suggest that the success of research alliance depends more on who is the partner than if the research alliance was made in the same or different ATC1, in which the licensor had previous successful experience in bringing a new drug to the market.

In the column 3, the variable “M&A” is split in “Vertical M&A” and “Horizontal M&A”. Both coefficients are statistically significant and negative with similar magnitude. If firms buy R&D projects in order to specialize or diversify to other ATC1, both strategies have similar innovation performance. However, when the variable “M&A” is split in “Small M&A” and “Big M&A” in the column 4 , the first coefficient is statistically significant and become more negative; whereas

Table 4.6: Probit Estimates for the Probability of Drug Approval

	(1)	(2)	(3)	(4)	(5)
Research Alliance			0.16**	0.17**	
			(0.07)	(0.07)	
M&A	-0.22**	-0.29***			
	(0.09)	(0.09)			
Vertical Alliance	0.26**				
	(0.11)				
Horizontal Alliance	0.27***				
	(0.08)				
Alliance_Small_Firm		-0.15*			-0.20***
		(0.09)			(0.08)
Alliance_Big_Firm		0.71***			0.57***
		(0.09)			(0.10)
Vertical M&A			-0.20**		
			(0.10)		
Horizontal M&A			-0.18*		
			(0.10)		
Small M&A				-0.41***	-0.41***
				(0.10)	(0.10)
Big M&A				-0.08	-0.07
				(0.10)	(0.10)
Biologic	0.33***	0.36***	0.32***	0.33***	0.34***
	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
Innovation Index	0.08	0.13	0.07	0.08	0.12
	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)
# of Diseases by Project	0.11***	0.11***	0.11***	0.11***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
University Project	-0.27*	-0.25*	-0.27*	-0.25*	-0.23*
	(0.14)	(0.13)	(0.14)	(0.14)	(0.13)
Vaccine	0.20	0.21	0.20	0.18	0.18
	(0.23)	(0.22)	(0.23)	(0.23)	(0.22)
Scale Firm	-0.08	-0.09	-0.08	-0.08	-0.09
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Scope Firm	-0.17***	-0.16***	-0.17***	-0.19***	-0.18***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Risk Portfolio	-3.44***	-3.49***	-3.38***	-3.46***	-3.48***
	(0.72)	(0.69)	(0.71)	(0.71)	(0.68)
Screening Ratio	0.25	0.29	0.27	0.27	0.30
	(0.22)	(0.23)	(0.22)	(0.22)	(0.22)
Firm Dummies	Yes	Yes	Yes	Yes	Yes
Disease Group	Yes	Yes	Yes	Yes	Yes
ATC1 (14 Dummies)	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Observations	14861	14861	14861	14861	14861

Note: Robust standard errors are in parentheses (clustered by firms).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: Vertical Alliance vs Horizontal Alliance

	Propensity Score Matching				
	ATET	Std. Errors	Z-Score	Lower CI	Upper CI
Vertical Alliance	-0.015	0.02	-0.68	-0.06	0.03
Observations	2112				

Note 1: Robust standard errors derived by Abadie and Imbens (2006 and 2012) are in parentheses.
Note 2: Covariates of Probit: Biologic, Innovation Index, # of Diseases by Project, University Project, Vaccine, Scale Firm, Scope Firm, Risk Portfolio, Screening Ratio, Disease, ATC1 and Year.
* p < 0.05, ** p < 0.01, and *** p < 0.001

the second coefficient is not statistically different than the in-house R&D projects, the reference category. This result suggests that R&D projects acquired from small-medium firms have lower probability of success than R&D projects acquired from big firms.

In column 5, the variables "Research Alliance" and "M&A" are split respectively in "Alliance_ Big_Firm" vs "Alliance_Small_Firm", and "Small M&A" vs "Big M&A". Regression 5 confirms the results of previous columns that size of partner matters for the innovation productivity, that is, R&D projects developed in alliance or acquired from Biotechs and medium firms on average have lower probability of drug approval than R&D projects developed in alliance or acquired from big pharmaceutical firms.

Table 4.7 presents the impact of "Vertical Alliance" against "Horizontal Alliance". The ATET is not statistically significant. This result confirms the conclusion of regression 1 in Table 4.6, that there is no difference in innovation performance between "Vertical Alliance" and "Horizontal Alliance". Table 4.8 presents the impact of "Alliance_Big_Firm" against "Alliance_Small_Firm". The ATET is statistically significant at 1%, that is, "Alliance_Big_Firm" has innovation performance about 10% higher than "Alliance_Small_Firm". This result is aligned to the conclusion of regressions 2 and 4 in Table 4.6. Table 4.9 presents the impact of "Vertical M&A" against "Horizontal M&A". The ATET is not statistically significant. This result confirms the conclusion of regression 3 in Table 4.6, that there is no difference in innovation performance between "Vertical M&A" and "Horizontal M&A". Table 4.10 presents the impact of "Big M&A" against "Small M&A". The ATET is statistically significant at 1%, that is, "Big M&A" has innovation performance about 3.5% higher than "Small M&A". This result is aligned to the conclusion of regressions 4 and 5 in Table 4.6.

Table 4.8: Alliance_Small_Firm vs Alliance_Big_Firm

	Propensity Score Matching				
	ATET	Std. Errors	Z-Score	Lower CI	Upper CI
Alliance_with_Big_Firm	0.107***	0.02	6.58	0.08	0.14
Observations	2112				

Note 1: Robust standard errors derived by Abadie and Imbens (2006 and 2012) are in parentheses.
Note 2: Covariates of Probit: Biologic, Innovation Index, # of Diseases by Project, University Project, Vaccine, Scale Firm, Scope Firm, Risk Portfolio, Screening Ratio, Disease, ATC1 and Year.
* p < 0.05, ** p < 0.01, and *** p < 0.001

Table 4.9: Vertical M&A vs Horizontal M&A

	Propensity Score Matching				
	ATET	Std. Errors	Z-Score	Lower CI	Upper CI
Vertical M&A	0.001	0.01	0.17	-0.01	0.02
Observations	6702				

Note 1: Robust standard errors derived by Abadie and Imbens (2006 and 2012) are in parentheses.
Note 2: Covariates of Probit: Biologic, Innovation Index, # of Diseases by Project, University Project, Vaccine, Scale Firm, Scope Firm, Risk Portfolio, Screening Ratio, Disease, ATC1 and Year.
* p < 0.05, ** p < 0.01, and *** p < 0.001

Table 4.10: Small M&A vs Big M&A

	Propensity Score Matching				
	ATET	Std. Errors	Z-Score	Lower CI	Upper CI
Big M&A	0.035***	0.01	6.86	0.03	0.05
Observations	6702				

Note 1: Robust standard errors derived by Abadie and Imbens (2006 and 2012) are in parentheses.
Note 2: Covariates of Probit: Biologic, Innovation Index, # of Diseases by Project, University Project, Vaccine, Scale Firm, Scope Firm, Risk Portfolio, Screening Ratio, Disease, ATC1 and Year.
* p < 0.05, ** p < 0.01, and *** p < 0.001

4.7 Conclusion

This study found that: 1) on average the R&D projects developed by research alliance have higher innovation performance than R&D projects developed internally, the exception is the R&D projects developed by research alliances with a small firms, in which the probability of drug approval is lower than R&D projects developed internally; 2) R&D projects originated from M&A have lower innovation performance than R&D projects developed internally; 3) R&D projects developed in research alliances with a big firm have higher probability to receive drug approval than R&D projects developed by research alliances with a small firm; 4) R&D projects acquired via big scale M&A have higher probability to receive drug approval than R&D projects acquired via small-medium scale M&A, and 5) there is no difference in innovation productivity between the strategies specialization and diversification at ATC level 1.

Together these findings suggest that big pharmaceutical firms are not in-licensing "lemons" projects or at least complementarities/synergies generated by the partnership more than compensated loss due transaction cost and asymmetric information problems. Although, on average it looks that R&D projects originated from small scale M&A are "lemons" projects, it might be the case that firms are buying technology (research material, research tool, or specific technique) to be integrated to internal capabilities and not an expected successful project. Furthermore, R&D projects developed in alliance have higher probability of drug approval than R&D projects from M&A, because the first is a specific selected combination of technology, know-how, research tools, and effort; whereas the second is an acquisition of a bundle of R&D projects not individually targeted, and consequently is more likely subject to asymmetric information problems, uncertainty, and deeper strategic problem such as a pipeline crisis.

Although we run several specifications, control for several factors, and check our results with matching estimators, our study has some limitations that can be addressed in future research. First, we treat each R&D project originated from M&A as an individual and independent R&D project. This assumption is more reasonable for internal R&D projects and R&D projects developed in alliance, but not for R&D projects acquired by bundles via M&A. Furthermore, the rationality of M&A goes beyond just improving innovation performance. Firms might engage in M&A to acquire market power, distribution channel, brand, scale in production and advertising, etc. Second, this

study was restricted to technical probability of success (drug approval), but little is known how the three strategies making, allying, and buying affects the commercial probability of success, such as revenue, market size, and probability to get a blockbuster. Third, although we pointed out to theories that justify our hypothesis, we didn't model the specific mechanisms that driver the differences in innovation performance among the strategies: making, allying, and buying.

CHAPTER 5

CONCLUSION

We investigated three aspects of drug development process little explored in the literature: 1) comparative performance between NPOs and firms in conducting clinical trials; 2) how scientific and technological specialization affects the timing of technological collaboration; and 3) how the strategies making, allying, and buying affect the probability of drug approval. The main findings of this study are: 1) firms complete each phase of clinical trial on average 13 months earlier than non-profit organizations; 2) an increase of 10% in the licensor scientific specialization speeds up the technological collaboration in 4 months, and an increase of 10% in the licensor technological specialization speeds up the technological collaboration in 5 months; and 3) considering that internal R&D projects have a probability of drug approval of 4.2%, research alliance has probability of drug approval 1.5% higher than internal R&D projects; whereas R&D projects originated from M&A has probability of drug approval 1.2% lower than internal R&D projects. Together these findings inform policy makers and managers on how to foster the innovation performance in the biopharmaceutical industry.

One limitation of this study is that we couldn't integrate in our analysis the cost of clinical trial, the cost of R&D project, and the cost of research alliance. Neither we use in our analysis the drug revenue, or other measure of commercial success like the probability of being blockbuster. If private data becomes available, we can develop further our study. Other limitation is that we estimated reduced form equations. For future research, we could model the specific mechanisms at organization level that firms use: 1) to complete clinical trials faster than NPOs; 2) to speed up the technological collaboration; and 3) to take decision among making, allying, and buying R&D.

An interesting research question related to our thesis is to investigate if patents and scientific

publications are complements or substitutes in terms of signaling strategies in the market for technology. Or in what conditions or factors make patents and scientific publications complements or substitutes. For example, firms and NPOs might use different signaling strategy related to patents and scientific publications. Our new methodology to track patents and scientific publications that cite specific technologies and diseases could be used for this purpose and be applied to other high tech sectors as well. Other research question to be explored given the availability of data is how firms might restrict the publication of clinical trial results in comparison to NPOs. It is well known that firms are more likely to impose some publication restriction to their employees, but there are a lot of heterogeneity among firms.

BIBLIOGRAPHY

Abadie, A., Imbens, G. W., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74 (1), 235–267.

Abadie, A., Imbens, G. W., 2012. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics.

Abadie, A., Imbens, G. W., 2016. Matching on the estimated propensity score. Econometrica 84 (2), 781–807.

Abrantes-Metz, R. M., Adams, C., Metz, A. D., 2004. Pharmaceutical development phases: a duration analysis. FTC Bureau of Economics working paper No. 274.

Adams, C. P., Brantner, V. V., 2006. Estimating the cost of new drug development: is it really $802 million? Health Affairs 25 (2), 420–428.

Aghion, P., Dewatripont, M., Stein, J. C., 2008. Academic freedom, private-sector focus, and the process of innovation. The RAND Journal of Economics 39 (3), 617–635.

Aghion, P., Tirole, J., 1994. The management of innovation. The Quarterly Journal of Economics, 1185–1209.

Akerlof, G., 1970. The market for lemons: qualitative uncertainly and market mechanism. The Quarterly Journal of Economics 89.

Allain, M.-L., Henry, E., Kyle, M., 2009. The timing of licensing: theory and empirics. Unpublished paper. Available from the authors.

Anton, J. J., Yao, D. A., 1994. Expropriation and inventions: Appropriable rents in the absence of property rights. The American Economic Review, 190–209.

Anton, J. J., Yao, D. A., 2002. The sale of ideas: Strategic disclosure, property rights, and contracting. The Review of Economic Studies 69 (3), 513–531.

Arora, A., Belenzon, S., Rios, L. A., 2014. Make, buy, organize: The interplay between research, external knowledge, and firm structure. Strategic Management Journal 35 (3), 317–337.

Arora, A., Gambardella, A., 1990. Complementarity and external linkages: the strategies of the large firms in biotechnology. The Journal of Industrial Economics 38 (4), 361–379.

Arora, A., Gambardella, A., 1994. Evaluating technological information and utilizing it: Scientific knowledge, technological capability, and external linkages in biotechnology. Journal of Economic Behavior & Organization 24 (1), 91–114.

Arora, A., Gambardella, A., 2010. The market for technology. Handbook of the Economics of Innovation 1, 641–678.

Arora, A., Gambardella, A., Magazzini, L., Pammolli, F., 2009. A breath of fresh air? firm type, scale, scope, and selection effects in drug development. Management Science 55 (10), 1638–1653.

Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In: The rate and direction of inventive activity: Economic and social factors. Princeton University Press, pp. 609–626.

Azoulay, P., 2002. Do pharmaceutical sales respond to scientific evidence? Journal of Economics & Management Strategy 11 (4), 551–594.

Azoulay, P., 2003. Agents of embeddedness. National Bureau of Economic Research (Working Paper 10142), 1–49.

Baker, D., 2008. The benefits and savings from publicly funded clinical trials of prescription drugs. International Journal of Health Services 38 (4), 731–750.

Berndt, E. R., Cockburn, I. M., 2014. Price indexes for clinical trial research: a feasibility study. Monthly Labor Review.

Berndt, E. R., Pindyck, R. S., Azoulay, P., 2003. Consumption externalities and diffusion in pharmaceutical markets: Antiulcer drugs. The Journal of Industrial Economics 51 (2), 243–270.

Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347–1393.

Budish, E., Roin, B. N., Williams, H., 2015. Do firms underinvest in long-term research? evidence from cancer clinical trials. The American Economic Review 105 (7), 2044–2085.

Cameron, A. C., Trivedi, P. K., 2005. Microeconometrics: methods and applications. Cambridge University Press.

Cassiman, B., Veugelers, R., 2006. In search of complementarity in innovation strategy: Internal r&d and external knowledge acquisition. Management Science 52 (1), 68–82.

Ceccagnoli, M., Higgins, M. J., Palermo, V., 2014. Behind the scenes: Sources of complementarity in r&d. Journal of Economics & Management Strategy 23 (1), 125–148.

Chandler, A. D., 2009. Shaping the industrial century: The remarkable story of the evolution of the modern chemical and pharmaceutical industries. Vol. 46. Harvard University Press.

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.

Cockburn, I. M., Henderson, R. M., 2000. Publicly funded science and the productivity of the pharmaceutical industry. Innovation Policy and the Economy 1, 1–34.

Cohen, W. M., Levinthal, D. A., 1989. Innovation and learning: the two faces of r & d. The Economic Journal 99 (397), 569–596.

Colombo, M. G., Garrone, P., 1996. Technological cooperative agreements and firm's r & d intensity. a note on causality relations. Research Policy 25 (6), 923–932.

Contractor, F. J., Ra, W., 2002. How knowledge attributes influence alliance governance choices: A theory development note. Journal of International Management 8 (1), 11–27.

Cox, D. R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological) 34 (2), 187–220.

CTTI, 2016. Points to consider for statistical analysis using the database for Aggregate Analysis of ClinicalTrials.gov. Clinical Trials Transformation Initiative.

Danzon, P. M., Nicholson, S., Pereira, N. S., 2005. Productivity in pharmaceutical–biotechnology r&d: the role of experience and alliances. Journal of Health Economics 24 (2), 317–339.

Das, T. K., Teng, B.-S., 2000. A resource-based theory of strategic alliances. Journal of Management 26 (1), 31–61.

Dehejia, R. H., Wahba, S., 2002. Propensity score-matching methods for nonexperimental causal studies. Review of Economics and Statistics 84 (1), 151–161.

Denicolo, V., 1996. Patent races and optimal patent breadth and length. The Journal of Industrial Economics, 249–265.

Dimaggio, P., Powell, W., 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review 48 (2), 147–160.

DiMasi, J. A., 2001. New drug development in the united states from 1963 to 1999. Clinical Pharmacology & Therapeutics 69 (5), 286–296.

DiMasi, J. A., Grabowski, H. G., 2007. The cost of biopharmaceutical r&d: is biotech different? Managerial and Decision Economics 28 (4-5), 469–479.

DiMasi, J. A., Hansen, R. W., Grabowski, H. G., 2003. The price of innovation: new estimates of drug development costs. Journal of health economics 22 (2), 151–185.

DiMasi, J. A., Hansen, R. W., Grabowski, H. G., Lasagna, L., 1991. Cost of innovation in the pharmaceutical industry. Journal of Health Economics 10 (2), 107–142.

DiMasi, J. A., Paquette, C., 2004. The economics of follow-on drug research and development. Pharmacoeconomics 22 (2), 1–14.

Dranove, D., Meltzer, D., 1994. Do important drugs reach the market sooner? The RAND Journal of Economics, 402–423.

Drennan, K. B., 2002. Patient recruitment: the costly and growing bottleneck in drug development. Drug Discovery Today 7 (3), 167–170.

Dubois, P., de Mouzon, O., Scott-Morton, F., Seabright, P., 2015. Market size and pharmaceutical innovation. The RAND Journal of Economics 46 (4), 844–871.

Fleischhacker, A., Ninh, A., Zhao, Y., 2015. Positioning inventory in clinical trial supply chains. Production and Operations Management 24 (6), 991–1011.

Fleming, T. R., DeMets, D. L., 1996. Surrogate end points in clinical trials: are we being misled? Annals of Internal Medicine 125 (7), 605–613.

Fosfuri, A., 2006. The licensing dilemma: understanding the determinants of the rate of technology licensing. Strategic Management Journal 27 (12), 1141–1158.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., Granger, C. B., 2015. Fundamentals of clinical trials. Springer.

Fudenberg, D., Gilbert, R., Stiglitz, J., Tirole, J., 1983. Preemption, leapfrogging and competition in patent races. European Economic Review 22 (1), 3–31.

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

Gambardella, A., 1995. Science and innovation: The US pharmaceutical industry during the 1980s. Cambridge University Press.

Gans, J. S., Hsu, D. H., Stern, S., 2008. The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. Management Science 54 (5), 982–997.

Garcia-Vega, M., 2006. Does technological diversification promote innovation?: An empirical analysis for european firms. Research Policy 35 (2), 230–246.

Getz, K. A., De Bruin, A., 2000. Breaking the development speed barrier: Assessing successful practices of the fastest drug development companies. Drug Information Journal 34 (3), 725–736.

Grabowski, H., Kyle, M., 2008. 11. Mergers and alliances in pharmaceuticals: effects on innovation and R&D productivity. Edward Elgar Publishing.

Griliches, Z., 1957. Hybrid corn: An exploration in the economics of technological change. Econometrica 25 (4), 501–522.

Grossman, S. J., Hart, O. D., 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. The Journal of Political Economy, 691–719.

Guedj, I., 2005. Ownership vs. contract: How vertical integration affects investment decisions in pharmaceutical r&d. McCombs Research Paper Series No. FIN-01-06.

Hart, O., 2003. Incomplete contracts and public ownership: remarks, and an application to public-private partnerships. The Economic Journal 113 (486), C69–C76.

Hart, O., Moore, J., 1990. Property rights and the nature of the firm. Journal of Political Economy, 1119–1158.

Heckman, J. J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In: Annals of Economic and Social Measurement, Volume 5, number 4. NBER, pp. 475–492.

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.

Henderson, R., Cockburn, I., 1996. Scale, scope, and spillovers: the determinants of research productivity in drug discovery. The Rand Journal of Economics, 32–59.

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.

Hoppe, H. C., 2002. The timing of new technology adoption: theoretical models and empirical evidence. The Manchester School 70 (1), 56–76.

Isakov, L., Lo, A. W., Montazerhodjat, V., 2016. Is the fda too conservative or too aggressive?: A bayesian decision analysis of clinical trial design. Draft.

Jayadev, A., Stiglitz, J., 2009. Two ideas to increase innovation and reduce pharmaceutical costs and prices. Health Affairs 28 (1), w165–w168.

Johnson, J. R., Williams, G., Pazdur, R., 2003. End points and united states food and drug administration approval of oncology drugs. Journal of Clinical Oncology 21 (7), 1404–1411.

Juni, P., Altman, D. G., Egger, M., 2001. Assessing the quality of controlled clinical trials. British Medical Journal 323 (7303), 42.

Kale, P., Singh, H., Perlmutter, H., 2000. Learning and protection of proprietary assets in strategic alliances: Building relational capital. Strategic Management Journal, 217–237.

Kamien, M. I., Schwartz, N. L., 1974. Patent life and r and d rivalry. The American Economic Review 64 (1), 183–187.

Katila, R., Mang, P. Y., 2003. Exploiting technological opportunities: the timing of collaborations. Research Policy 32 (2), 317–332.

Koenig, M. E., 1983. A bibliometric analysis of pharmaceutical research. Research Policy 12 (1), 15–36.

Lacetera, N., 2009. Different missions and commitment power in r&d organizations: Theory and evidence on industry-university alliances. Organization Science 20 (3), 565–582.

Lerner, J., Shane, H., Tsai, A., 2003. Do equity financing cycles matter? evidence from biotechnology alliances. Journal of Financial Economics 67 (3), 411–446.

Lewis, T., Reichman, J., So, A., et al., 2007. The case for public funding and public oversight of clinical trials. Economists' Voice 4 (1), 1–4.

Lichtenberg, F., Philipson, T., 2002. The dual effects of intellectual property regulations: Within? and between? patent competition in the u.s. pharmaceuticals industry. The Journal of Law & Economics 45 (S2), 643–672.

Lichtenberg, F. R., 2005. The impact of new drug launches on longevity: evidence from longitudinal, disease-level data from 52 countries, 1982–2001. International journal of health care finance and economics 5 (1), 47–73.

Loftus, P., 2015. Drug makers buy pricey vouchers to speed products to market. Wall Street Journal.

Malerba, F., Orsenigo, L., 2015. The evolution of the pharmaceutical industry. Business History 57 (5), 664–687.

McMillan, G. S., Narin, F., Deeds, D. L., 2000. An analysis of the critical role of public science in innovation: the case of biotechnology. Research Policy 29 (1), 1–8.

Merck, 2014. 2013 FORM 10-K - U.S. Securities and Exchange Commission. Merck & Co., Inc.

Merton, R. K., 1973. The sociology of science: Theoretical and empirical investigations. University of Chicago Press.

Miller, D. J., 2006. Technological diversity, related diversification, and firm performance. Strategic Management Journal 27 (7), 601–619.

Mirowski, P., Van Horn, R., 2005. The contract research organization and the commercialization of scientific research. Social Studies of Science 35 (4), 503–548.

Mittra, J., 2006. The socio-political economy of pharmaceutical mergers: A case study of sanofi and aventis. Technology Analysis & Strategic Management 18 (5), 473–496.

Mittra, J., 2007. Life science innovation and the restructuring of the pharmaceutical industry: Merger, acquisition and strategic alliance behaviour of large firms. Technology Analysis & Strategic Management 19 (3), 279–301.

Nelson, R. R., 1959. The simple economics of basic scientific research. The Journal of Political Economy 67 (3), 297–306.

Nelson, R. R., Winter, S. G., 1982. An evolutionary theory of economic change. Harvard University Press.

Nesta, L., Saviotti, P. P., 2005. Coherence of the knowledge base and the firm's innovative performance: evidence from the us pharmaceutical industry. The Journal of Industrial Economics 53 (1), 123–142.

NIH, 2015. NIH Clinical Center Director's Annual Report 2015. National Institutes of Health.

Oxley, J. E., 1997. Appropriability hazards and governance in strategic alliances: A transaction cost approach. Journal of Law, Economics, and Organization 13 (2), 387–409.

Partha, D., David, P. A., 1994. Toward a new economics of science. Research Policy 23 (5), 487–521.

Pfizer, 2014. 2013 FORM 10-K - U.S. Securities and Exchange Commission. Pfizer Inc.

Pisano, G. P., 1989. Using equity participation to support exchange: Evidence from the biotechnology industry. Journal of Law, Economics, & Organization 5 (1), 109–126.

Pisano, G. P., 1997. R&d performance, collaborative arrangements and the market for know-how: A test of the hypothesis in biotechnology. SSRN Working Paper.

Polanyi, M., 2000. The republic of science: Its political and economic theory. Minerva 38 (1), 1–21.

Puhani, P., 2000. The heckman correction for sample selection and its critique. Journal of Economic Surveys 14 (1), 53–68.

Quintana-García, C., Benavides-Velasco, C. A., 2008. Innovative competence, exploration and exploitation: The influence of technological diversification. Research Policy 37 (3), 492–507.

QuintilesIMS, 2016. QuintilesIMS Institute Forecast: Global Drug Market Will Reach Nearly $1.5 Trillion in 2021 as Spending Growth Moderates. QuintilesIMS Institute.

Rafols, I., Hopkins, M. M., Hoekman, J., Siepel, J., O'Hare, A., Perianes-Rodríguez, A., Nightingale, P., 2014. Big pharma, little science?: A bibliometric perspective on big pharma's r&d decline. Technological Forecasting and Social Change 81, 22–38.

Reinganum, J. F., 1989. The timing of innovation: Research, development, and diffusion. Handbook of Industrial Organization 1, 849–908.

Roland, G., Xu, J., Wagner, J., 2015. First-to-patent does not predict first-or best-in-class: Analysis of approved small molecule vs. biologic drugs. Clinical Pharmacology & Therapeutics 97 (1), 19–21.

Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika, 41–55.

Rothaermel, F. T., 2001. Complementary assets, strategic alliances, and the incumbent's advantage: an empirical study of industry and firm effects in the biopharmaceutical industry. Research policy 30 (8), 1235–1251.

Rothaermel, F. T., Deeds, D. L., 2004. Exploration and exploitation alliances in biotechnology: A system of new product development. Strategic management journal 25 (3), 201–221.

Santoro, M. D., McGill, J. P., 2005. The effect of uncertainty and asset co-specialization on governance in biotechnology alliances. Strategic Management Journal 26 (13), 1261–1269.

Sauermann, H., Stephan, P., 2013. Conflicting logics? a multidimensional view of industrial and academic science. Organization Science 24 (3), 889–909.

Scherer, F. M., 2010. Pharmaceutical innovation. Handbook of the Economics of Innovation 1, 539–574.

Schulze, U., Ringel, M., 2013. What matters most in commercial success: first-in-class or best-in-class? Nature Reviews Drug Discovery 12 (6), 419–420.

Schwartzman, D., 1976. Innovation in the pharmaceutical industry. Johns Hopkins University Press.

Stern, S., 2004. Do scientists pay to be scientists? Management Science 50 (6), 835–853.

Stoneman, P., Battisti, G., 2010. Chapter 17 - the diffusion of new technology. In: Hall, B. H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation, Volume 2. Vol. 2. North-Holland, pp. 733 – 760.

Stuart, T. E., Ozdemir, S. Z., Ding, W. W., 2007. Vertical alliance networks: The case of university–biotechnology–pharmaceutical alliance chains. Research Policy 36 (4), 477–498.

Toh, P. K., Kim, T., 2013. Why put all your eggs in one basket? a competition-based view of how technological uncertainty affects a firm's technological specialization. Organization Science 24 (4), 1214–1236.

Trusheim, M. R., Aitken, M. L., Berndt, E. R., 2010. Characterizing markets for biopharmaceutical innovations: do biologics differ from small molecules? In: Forum for Health Economics & Policy. Vol. 13.

Wagner, J., Williams, S., Webster, C., 2007. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. Clinical Pharmacology & Therapeutics 81 (1), 104–107.

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

Winslow, R., Walker, J., 2014. Drug firms buy $67.5 million voucher to speed fda review. Wall Street Journal.

Wooldridge, J. M., 2010. Econometric analysis of cross section and panel data. MIT press.

Zucker, L. G., Darby, M. R., Armstrong, J. S., 2002. Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. Management Science 48 (1), 138–153.